INSTITUTE FOR DEFENSE ANALYSES

Comparative Investigation of Source Term Estimation Algorithms for Hazardous Material Atmospheric Transport and Dispersion Prediction Tools

N. Platt, Project Leader
D. F. DeRiggi

July 2012

IDA Document D-4048
Log: H 10-000243

Copy

About this Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract DASW01-04-C-0003, Tasks DC-1-2607 and DC-1-2615, "Support for the Defense threat Reduction Agency (DTRA) in the Validation Analysis of Hazardous Material Assessment Model, Support for DTRA in the Validation Analysis of Hazardous Material Transport and Dispersion Prediction Models." The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

The authors would like to thank the IDA committee, Dr. Steve Warner (Chair), Dr. James F. Heagy, Dr. David Spalding, and Dr. Jeffry T. Urban for providing technical review of this effort.

I N S T I T U T E   F O R   D E F E N S E   A N A L Y S E S

IDA Document D-4048

Comparative Investigation of Source Term Estimation Algorithms for Hazardous Material Atmospheric Transport and Dispersion Prediction Tools

N. Platt, Project Leader
D. F. DeRiggi

Executive Summary

A field experiment was conducted in 2007 during which a tracer gas was released into the atmosphere and its dispersal was tracked on a dense grid of samplers. The goal of this field trial was to provide information to further the development of source term estimation (STE) algorithms capable of predicting release location and characteristics (e.g., time of release and amount of material released). After the field trial, several algorithm developers participated in an exercise in which they provided protocol-controlled – and hence comparable – predictions of the release source characteristics based on select data collected during the experiment. The goal of this document is to describe the results of our assessments and to compare these algorithms based on their protocol-controlled predictions. This analysis is meant to help the Department of Defense (DoD) identify the current state of STE algorithm development (identify the "state of the art"), and it provides specific and constructive feedback to participating STE developers.

In September 2007 at the U.S. Army's Dugway Proving Ground, a short-range atmospheric dispersion field experiment called the Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07) was conducted. FFT 07 was designed to collect information to support the development of prototype STE algorithms to back-predict the source(s) of a hazardous materials release when given detection data from sensors and local meteorological conditions. A total of 82 trials, involving a mix of instantaneous and continuous releases from up to four simultaneous sources of a neutrally buoyant tracer gas (propylene), were conducted over a period of 2½ weeks. These releases occurred during both daytime and at night. The tracer gas was sampled on a dense regular grid of samplers approximately 450 meters by 450 meters.

A comparative investigation of STE algorithms based on this field experiment began in 2008. Participating algorithm developers were asked to predict the source of a tracer gas release based on limited information from the tracer measurements and local meteorological conditions. Depending on the individual algorithms' capabilities, they were tasked to predict the location of the sources of the release, the number of sources, the mass of each source, and the timing of the release from each source.

The general method of this investigation was first to provide participating developers with a subset of sensor data that was collected during selected FFT 07 releases – individual cases were constructed from the subset of FFT 07 releases for which source term predictions were sought. However, they were not provided with any information (e.g., time, location, or mass) on the actual source release that they were asked to predict – that is, these were "blind" predictions. Since a partial set of the original field trial data (including source term information) was released to participants to help develop algorithms, some tracer and meteorological data were concealed so that algorithm developers would not be able to easily determine which physical FFT 07 release was used to create a particular case.

Next, algorithm developers provided "blind" predictions that could then be compared to parameters of the actual release. This investigation consisted of 104 individual cases of sensor data that were distributed in September 2008. These cases provided high-resolution, continuous streams of concentration data for ingestion by STE algorithms. The complexity and degree of the information provided in individual cases were varied in that the algorithms were sometimes asked to predict cases in which:

- The meteorology was relatively well-characterized and detection data were available from a relatively large number of chemical sensors in order to characterize STE algorithm performance under optimistic conditions.
- The meteorological data and number of available sensors were more limited in order to characterize STE algorithm performance under less ideal, but perhaps more realistic, conditions.

A total of 8 STE algorithm developers participated in this investigation providing a total of 14 full and partial sets of predictions. Some developers provided multiple predictions based on different algorithms under development. We particularly note that not all developers submitted predictions for all 104 cases. Some algorithms were not capable of predicting certain types of releases that were considered (e.g., instantaneous or continuous).[1] Some model developers selectively limited their predictions to cases when a relatively large number of sensors (e.g., 16) were provided, or, because of funding and timing constraints, limited their set of predictions to either the first "53" or some "semi-random" subset of cases.

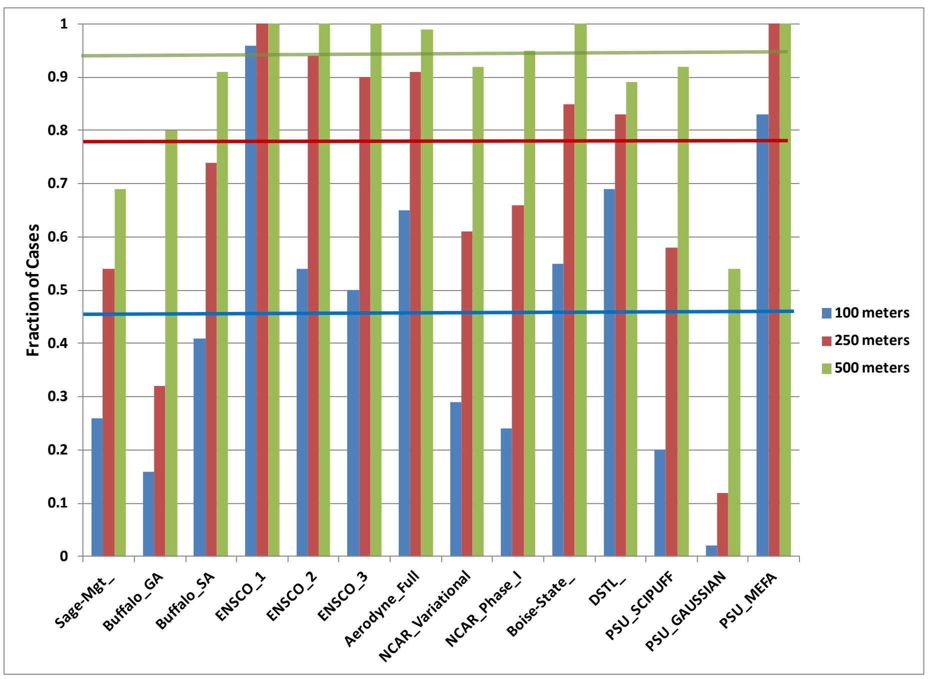
The goal of these evaluations was not to declare a "winning" algorithm, but rather to assess the state of the art in the area of source term estimation and provide constructive feedback to the developers. Therefore, we started our analysis by evaluating algorithm performance trends instead of analyzing each individual algorithm. We did not attempt to determine whether the predictions were "good enough" for a particular operational use. Two separate methodologies were pursued: (1) comparison of selected top-level algorithm performance metrics under a variety of conditions and among algorithms and (2) application of linear regression techniques to discern trends among different algorithms. Two top-level performance metrics were constructed to compare STE algorithm performance. For each individual case predicted by an STE algorithm, two

---

[1] In this case, algorithm developers tried to selectively prescreen tracer information to ascertain whether a particular release fell within a selected class.

measures were calculated: (1) the distance between the average predicted and the average observed location of the source(s) that we refer to as "miss distance"[2] and (2) the ratio of the total predicted mass to the total released mass from all sources that we refer to as "mass ratio."

The following figure shows results of these calculations for "miss distance" at three levels of interest. For each set of STE predictions, the grouped bars denote the fraction of predictions that are less than the particular level of interest. With respect to our miss distance metric, all algorithms were able to predict "averaged" source term locations to within 500 meters (i.e., a size comparable to the size of the tracer measurement grid of the FFT 07 experiment), and a wide variation in the quality of the algorithm predictions was seen when the miss distance was on the order of tens of meters (i.e., less than 100 meters). Few algorithms are able to consistently predict the source of a release with an accuracy of more than a few hundred meters. We note that the FFT 07 sensor grid was less than approximately 500 meters across and that the release sources were less than 100 meters away from the leading edge of the sensor grid.



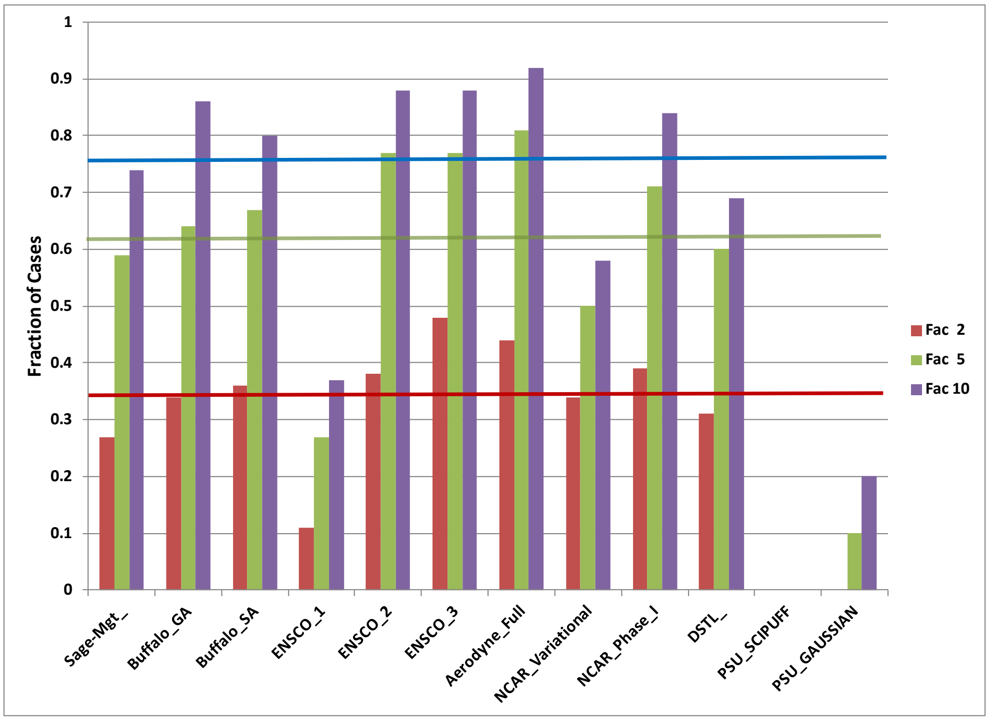
Horizontal lines correspond to medians of fractions for all algorithms and at various thresholds: 0.46 (blue line) for the fraction of miss distances less than 100 meters, 0.79 (brown line) for the fraction of miss distances less than 250 meters, and 0.94 (green line) for the fraction of miss distances less than 500 meters. Therefore, these lines separate the algorithms into better and worse performing halves, as measured by the given metric calculated over all cases for each algorithm.

Algorithm Inter-Comparison Using Averaged Miss Distance Fraction of Cases below 100, 200, and 500 meters

---

2 The distance between the predicted and observed location for an individual source can be larger or smaller than the miss distance metric value that corresponds to an average difference when more than one location is involved in the release or prediction.

With respect to predicting release mass, algorithm performance varied widely. For each set of predictions, the following figure shows the fractions of cases in which observed and predicted masses were within factors of 2, 5, and 10 of each other. About half of the models were able to predict total mass of the source to within a factor of 10 for about three-quarters of the cases. When the prediction standard quality was raised to within a factor of 2, about half of the algorithms had this level of accuracy for less than one-third of the cases. Most evaluated STE algorithms did not consistently predict total mass to within a factor of 5. We caution that these results capture global algorithm performance without any effort to ensure that compared predictions are compatible with each other. For instance, as noted earlier, these results do not take into account that some algorithms provided only partial predictions (i.e., not a complete set of predictions for all cases), and some algorithm developers picked preferred sets of predictions to submit.



Thick colored lines correspond to the medians of the fractions for all of the algorithms and at the various thresholds: 0.34 (brown line) for factor of 2, 0.62 (green line) for factor of 5, and 0.77 (blue line) for factor of 10.

Algorithm Inter-Comparison Using Observed and Predicted Mass Fractions within Factors of 2, 5, and 10 of Each Other

Our analyses that applied linear regression techniques indicated that the time of the release (night versus day), type of meteorology provided (detailed versus sparse "operational"), and the number of simulated sensors (4 versus 16) *did not* lead to significant differences in prediction quality for most of the STE algorithms under evaluation. At first glance, these results seem to be counterintuitive. For instance, one expects that quadrupling the number of sensors from 4 to 16, or using high-frequency close-in meteorology, should necessarily lead to better predictions. Also, time of release should, in general, be strongly correlated with the atmospheric stability and this should significantly affect atmospheric dispersion and could be hypothesized to affect (and perhaps differentially affect) STE algorithm performance. Thus, it is, at least at first glance, unexpected that the STE algorithms are capable of predicting source term parameters with equal skill under stable and unstable atmospheric conditions, or without regard to the number of sensors providing information on the tracer gas or the expected quality of the meteorological information. We hypothesize that the relatively small spatial scale of the FFT 07 sensor grid (approximately 450 by 450 meters), and the proximity of the release locations, both to each other and to the upwind leading edge of the sensor grid, might be responsible for these findings. For instance, for most single-source releases, the cross-wind extent of the plume does not cover more than a few neighboring sensors, and no significant spatial variation occurs in the plume over the sensor grid as the downwind distance from the release location increases. Thus, changing the number of simulated sensors from 4 to 16 might not provide enough additional information for the STE algorithms.

In addition, linear regression analysis indicated that the number of sources and the type of release [continuous release versus single realization of instantaneous puff(s) versus multiple realizations of instantaneous puff(s)] *are* significant variables in terms of predicting algorithm performance for most participating algorithms. We note that regression analysis itself (as we used it) does not quantify the quality of the algorithms' ability to predict source term parameters – it only indicates which release factors have an effect on the quality of the STE predictions.

Our most significant observations and recommendations from these investigations are described below:

- **Source term estimation, as envisioned for chemical and biological weapon attacks, remains a challenge.** An initial look at state-of-the-art STE algorithms participating in this exercise revealed shortcomings with respect to estimating spatial location and mass of the release. Although most STE algorithms were capable of estimating release location on a scale comparable to the limited size of the sensor grid used in FFT 07, and noting that the releases were very close to the upwind edge of the sensor grid, questions remain as to how well these algorithms would perform in operationally relevant scenarios that would undoubtedly include sensors spaced farther apart from each other and the release location.[3]

[3] One could conceive of an STE algorithm that places the source at the location of the first sensor that detects the release. This type of algorithm would be consistent with placing an Allied Tactical Publication-45 (ATP-45) warning triangle at the sensor that registers first detection. Given the limitations associated with the FFT 07 field experiment (especially the scale), such an algorithm would perform quite comparably to the more complex STE algorithms that were investigated.

- **The FFT 07 field trials appear to have limited applicability to practical validation of STE algorithms.** FFT 07 is the most comprehensive field experiment conducted to provide information to further the development and assessment of STE algorithms – certainly a valuable and necessary source of measurements and observations for this goal of improving the state of the art.[4] However, the relatively small size of the sensor grid and the closeness of the release locations to the upwind leading edge of the sensor grid, limit the usefulness of FFT 07 as the basis for future validation of an STE algorithm for militarily relevant scenarios. Moreover, our analysis revealed that certain input variations for the STE algorithms (such as quadrupling the number of available sensors or providing detailed high-resolution meteorology near the center of the sensor grid) did not lead to expected discernible improvements in the quality of STE predictions. This suggests that the small scale of FFT 07 – a few hundred meters – limited its usefulness for evaluations of even fundamental STE algorithm performance at larger (and for many applications, more realistic) scales where atmospheric stability, the quality of meteorological inputs, and the amount of available sensor (i.e., "detector") information can reasonably be hypothesized to influence STE algorithm performance.
- **A relatively high-fidelity, virtual, simulated environment could be useful for future assessments and independent validation activities of STE algorithms.** This recommendation rests on the premise that a relatively large-scale, realistic field trial is unaffordable (and possibly not executable in any case). As computational power becomes more available and relatively cheap, the potential exists to use computer modeling tools to supplement field testing of system components. The use of such tools holds the promise of increasing the efficiency of the conducted field tests, aiding the evaluation of results obtained from such tests, and reducing costs. We recommend that simulated environments such as the National Center for Atmospheric Research Virtual THreat Response Emulation and Analysis Testbed modeling system should be considered and take center stage to supplement and extend field trial data. Furthermore, if future assessment and validation efforts of STE modules will largely rely on simulated environments, future laboratory measurements or field trial designs and observations must take this into account. That is, we recommend a holistic approach to designing the strategy by which simulated environments and field trials (or laboratory tests) are used to further the assessment and validation of STE mod-

[4] The provided FFT 07 data were valuable to algorithm developers, especially in terms of refining their expectations. For instance, several prototype algorithms did not expect that (1) some sensors would have "noise" floors (i.e., they register some "signals" even when no tracer gas was present), and (2) different sensors have differing levels of noise. That necessitated some developers to implement new threshold algorithms before supplying the provided data to their algorithms.

ules. Such an approach should ensure future activities are complementary and should especially seek synergistic activities (e.g., field trial or laboratory observations that support increased confidence in aspects of the virtual environment that are critical to its use when applied to STE algorithm assessment and validation).

(This page is intentionally blank.)

Contents

A. Comparison of Algorithms Based on Averaged Miss Distance 7
B. Linear Regression Analysis Results 10
C. Comparison of Selected Global Algorithm Performance Metrics 14
D. Discussion 17
Appendix A. Glossary A-1
Appendix B. Sequence of Events B-1
Appendix C. Brief Description of Source Term Estimation Algorithms C-1
Appendix D. Developer Feedback Package Description D-1
Appendix E. Additional Plots for Miss Distance Inter-comparison E-1
Appendix F. Linear Regression Description F-1
Appendix G. "Cross-Term" Regression Results Tables G-1
Appendix H. Task Order Extract H-1

Figures

1. Schematic Lay-Down of the Subset of Instrumentation Used during FFT 07 Field Trials 2
2. Example of the Distance Metric Computation and Total Mass Calculation used to Compare Algorithm Performance for each Individual Case 7
3. Median "Miss" Distance for Individual STE Algorithms....................9
4. Algorithm Inter-Comparison Using Averaged Miss Distance Fraction of Cases below 100, 200, and 500 Meters....................15
5. Total Mass Over-Prediction Fraction for the 12 STE Algorithms that Provided Enough Information to Calculate Total Predicted Release Mass from All Sources.16
6. Algorithm Inter-Comparison Using Total Observed and Predicted Mass Fractions within Factors of 2, 5, and 10 of Each Other....................17

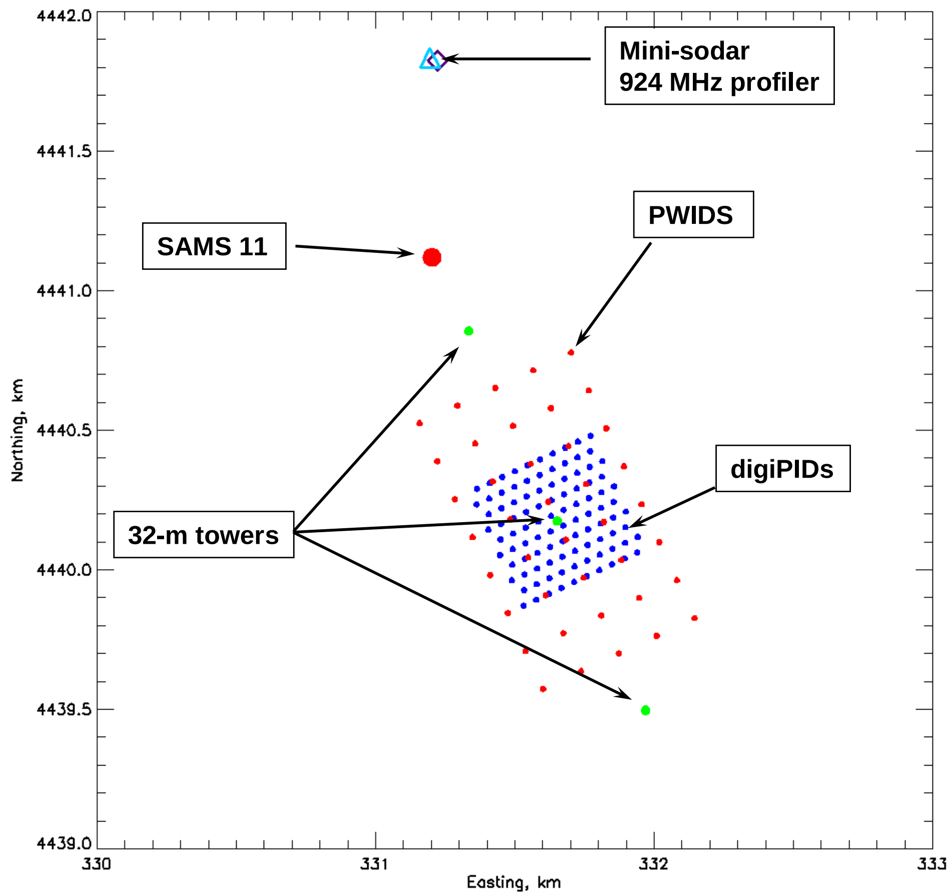
Tables

1. Composition of Cases Distributed to STE Algorithm Developers to Provide Predictions 3
2. Organizations That Participated in the Evaluation[a]....................5
3. Basic Capabilities of Each of the STE....................6
4. Composition of Single and Double Source Predicted Cases Provided by STE Developers 8
5. Table of Significant Factors for Backward Regression....................12
6. Table of Significant Factors for Stepwise Regression....................13

Comparative Investigation of Source Term Estimation Algorithms for Hazardous Material Atmospheric Transport and Dispersion Prediction Tools

When only a few sensors detect hazardous materials resulting in a warning, rapid provision of an estimate of the source location, time of release, and amount of released material is useful. Such an estimate can help refine predictions of the area affected by the release and can support near-term follow-on actions to investigate the cause and nature of the release. In some cases, refined predictions resulting from such source term estimation (STE) can support tactical decisions (e.g., which roads to travel on and which to avoid). For longer range situations (tens of kilometers), accurate estimates of source location can facilitate improved hazard-area predictions to better support warnings and possible evacuation, advocate the use of efficient mission-oriented protective posture gear, or perhaps enhance medical response. The Joint Effects Model (JEM), under acquisition through the Joint Program Office for Chemical and Biological Defense, is the DoD-wide transport and dispersion model intended to satisfy DoD requirements for chemical, biological, radiological, and nuclear (CBRN) hazard predictions and consequence assessment. The future JEM release (JEM 2.0) has an STE requirement that is yet to be satisfied. The Defense Threat Reduction Agency – Joint Science and Technology Office (DTRA-JSTO) has primary responsibility for science and technology development of JEM and is responsible for supplying JEM with this capability.

In September 2007, DTRA conducted a short-range (~500 meters), highly instrumented atmospheric transport and dispersion test at the U.S. Army's Dugway Proving Ground (DPG) [1]. This test, referred to as Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07), was designed to collect data to support further development of prototype algorithms. FFT 07 was sponsored by DTRA-JSTO for Chemical and Biological Defense (CBD) and was conceived of and planned within the Technical Panel 9 for Hazard Assessment (TP9) of The Technical Cooperation Program Chemical, Biological, and Radiological Defense group, thus making this effort an international collaboration (in this case, among the United States, United Kingdom, Canada, and Australia). A total of 82 trials involving a mix of instantaneous and continuous releases of a neutrally buoyant tracer gas (propylene) were conducted over a period of 2½ weeks. Tracer gas concentrations were measured on a dense regular grid of samplers approximately 450 meters by 450 meters. Figure 1 illustrates the layout of a subset of FFT 07 instrumentation including the 100 digital photo-ionization detectors (digiPIDs) that were used to sample propylene concentration at 50 Hz and the locations of various instruments that collected meteorological observations. Not shown in this schematic are 20 ultraviolet ion collector (UVIC) detectors positioned between the digiPIDs at lines 3 and 8.



Blue dots denote locations of 100 digiPIDs used to sample propylene concentrations at 50 Hz, small red dots denote locations of 40 Portable Weather Information and Display Systems (PWIDs) used to collect detailed surface meteorology, green dots denote locations of three 32-meter towers that carried additional meteorological instrumentation, large red dot denotes location of SAMS 11 meteorological weather station, and the diamond and triangle at the top denote location of mini-sodar and 924 MHz wind profiler.

Figure 1. Schematic Lay-Down of the Subset of Instrumentation Used during FFT 07 Field Trials

With respect to STE algorithm development, there were several reasons for conducting FFT 07. First, the field trial experiment was intended to provide a set of data that STE algorithm developers could use to improve their algorithms. Next, the collected information could be used to assist in identifying strengths and weaknesses of different modeling approaches chosen by developers.[1] Finally, the assessment of STE algorithms using data collected during FFT 07 was meant to help DoD identify the current state of the STE algorithms in general (the “state of the art”).

A comparative investigation of STE algorithms based on FFT 07 data began in late 2008. Appendix B shows the sequence of events that were part of this investigation and preceded this report. The general method of this research was to first provide participating developers with a subset of sensor data that was collected on selected FFT 07 trials. Next, developers provided “blind” predictions (e.g., algorithm developers did not know which physical trial corresponded to a particular case for which they were providing predictions) that were compared to parameters of the actual release. This investigation consisted of 104 individual cases of sensor data constructed from a subset of the available digiPID data that were distributed in September 2008. Table 1 list the composition of the cases. These cases included continuous streams of concentration data (1 Hz) for ingestion by STE algorithms.

Table 1. Composition of Cases Distributed to STE Algorithm Developers to Provide Predictions

Condition	All Trials	Single	Double	Triple	Quad
None	104	40	40	16	8
Puff	52	20	20	8	4
Cont	52	20	20	8	4
Daytime	52	20	20	8	4
Nighttime	52	20	20	8	4
Daytime/Puff	26	10	10	4	2
Daytime/Cont	26	10	10	4	2
Nighttime/Puff	26	10	10	4	2
Nighttime/Cont	26	10	10	4	2

This evaluation consisted of cases that equally sampled parameters that were expected to most significantly affect the quality of STE predictions. These parameters included the time of day of the tracer release (day or night), the type of tracer release (continuous or instantaneous – sometimes referred as “puff”), and the number of sensors reporting data (4 or 16). To provide some realism with respect to meteorological inputs, for some cases, developers were provided with surface wind velocity observations and a vertical wind velocity profile from sites up to 2 km removed from the tracer releases and

[1] This report does not deal with identifying strengths and weaknesses of individual algorithms and modeling approaches. It is left to individual STE algorithm developers to evaluate their algorithms. We exerted significant efforts in providing detailed (quantitative) feedback to STE algorithm developers in the forms of recurrent briefings and developer feedback packages, the contents of which are described in Appendix D.

sampler grid, instead of the more detailed meteorological observations that were made at the center location of the sampler grid. An additional sampled parameter, which could affect the quality of STE predictions, was the number of sources (single, double, triple, or quad). In cases of multiple sources, all individual sources were synchronized together (e.g., all air cannons were fired simultaneously for instantaneous releases, and all valves were tuned on/off at the same time for continuous releases). FFT 07 individual puff trials involved multiple (up to 10) realizations. These puffs were released by firing air cannons every few minutes resulting in “trains of puffs” periodically traversing the digiPID grid. Hence, for puff releases, some distributed cases included a single realization of the puff(s), but some of the distributed cases included multiple (up to 10) realizations. The full structure of the distributed cases including methodology to create individual cases is given in References 2 and 3.

A total of 8 different STE algorithm developers participated in this exercise. A total of 14 full and partial sets of predictions were received with some exercise participants providing multiple sets of predictions based on different algorithms that they have been developing. We note that not all algorithm developers submitted predictions for all 104 cases. Some algorithms were not capable of predicting certain types of the considered releases (e.g., instantaneous or continuous).[2] Some model developers selectively limited their predictions to cases when high numbers of simulated sensors (e.g., 16) were provided or, because of funding and timing constraints, limited their set of predictions to either the first “53” or some “semi-random” subset of the cases. Table 2 depicts the organizations that participated in the evaluation together with the composition of predicted cases that they provided.

[2] In this case, algorithm developers tried to selectively prescreen the tracer information to ascertain whether a particular release fell within a selected class.

Table 2. Organizations That Participated in the Evaluation and Composition of the Prediction Sets Received

Organization	Total	Cont	Puff	Daytime	Nighttime	Single	Double	Triple	Quad
Aerodyne	104	52	52	52	52	40	40	16	8
Boise-State	33	14	19	21	12	13	13	4	3
Buffalo/GA	104	52	52	52	52	40	40	16	8
Buffalo/SA	70	34	36	34	36	26	26	12	6
DSTL	35	5	30	20	15	12	14	7	2
ENSCO/Set 1	102	51	51	50	52	39	39	16	8
ENSCO/Set 2	104	52	52	52	52	40	40	16	8
ENSCO/Set 3	42	24	18	19	23	13	15	10	4
NCAR/ Variational	38	3	35	20	18	16	14	4	4
NCAR/Phase I	38	3	35	20	18	16	14	4	4
Sage-Mgt	104	52	52	52	52	40	40	16	8
PSU/Gaussian	50	26	24	25	25	18	20	8	4
PSU/SCIPUFF	50	26	24	25	25	18	20	8	4
PSU/MEFA	35	19	16	17	18	13	16	5	1

Composition of predicted cases that were provided are broken down into several categories including release type, time of day, and number of sources. The red font values denote that a full set of predictions was provided; blue font values denote that the predictions were provided for at least 50 percent of the distributed cases.

Table 3 lists some basic capabilities of each of the STE algorithms including their ability to predict the number (e.g., single, double, triple, or quad source) and types of sources (e.g., continuous or instantaneous puff release). The table also identifies the number of cases provided in the final set of predictions; the number of updates to, and replacements of, predictions provided; and a few additional comments. Appendix C has short technical description of each of the algorithms that participated in this investigation.

The goal of these evaluations was not to declare a "winning" algorithm, but rather to try to assess the state of the art in the area of source term estimation. We focused our analysis on the evaluation of algorithm performance trends, rather than analyzing each individual algorithm's performance. The developer feedback package that was distributed in September 2009 provided information pertaining to performance of individual algorithms. Appendix D describes the content of the developer feedback package and provides some sample plots. Individual STE algorithm developers should find this information useful for analyzing their algorithm's performance, perhaps finding areas for improvement, and eventually publishing their results.

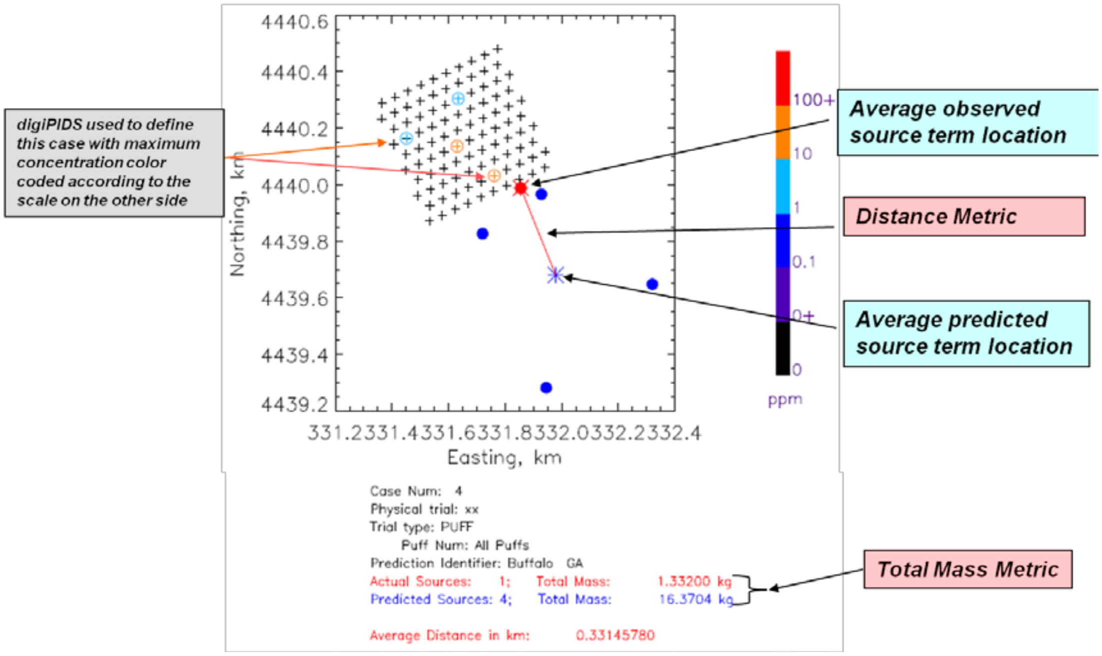
Table 3. Basic Capabilities of Each STE

Organization	Number of Sources	Type	Total Predicted Cases	Number of Updates and Comments
Aerodyne	Multi	Cont/Puff	104	Partial Set, Full Set
Boise-State	Single	Cont/Puff	33	First 30 case, First 53 cases
Buffalo/GA	Multi	Cont/Puff	104	
Buffalo/SA	Mostly Single	Cont/Puff	70	
DSTL	Single	Puff	35	
ENSCO/Set 1	Multi	Cont/Puff	102	
ENSCO/Set 2	Single	Cont	104	
ENSCO/Set 3	Single	Cont	42	Set 3 is a subset of Set 2 that uses larger search box
NCAR/Variational	Single	Puff	38	3 updates/replacements
NCAR/Phase I	Single	Puff	38	3 updates/replacements
Sage-Mgt	Single	Cont/Puff	104	3 updates/replacements
PSU/Gaussian	Single	Cont/Puff	50	16 sensor cases only
PSU/SCIPUFF	Single	Cont/Puff	50	16 sensor cases only
PSU/MEFA	Multi	Cont/Puff	35	16 sensor cases only

Only predicted cases that contain source term location are counted in this table. Some algorithm developers provided predictions for cases that did not converge to a particular location but did estimate other source term parameters such as type or mass of the release. For instance, Boise-State provided predictions for the first 53 cases, but only 33 of these reported a particular location.

As depicted in Table 3, individual STE algorithm developers who participated in the evaluations have different capabilities with respect to predicting the numbers and types of sources. In order to fairly compare these algorithms, we needed to define common metrics applicable to all algorithms. We selected two metrics: the distance between the averaged predicted and averaged observed source term locations and the ratio of the observed to predicted release mass from all sources. Figure 2 illustrates the distance metric calculation.[3] From the 14 sets of STE algorithm predictions, 12 algorithms provided enough information to calculate the mass ratio, and all 14 provided enough data to calculate our distance metric.

[3] Given the varying capabilities of individual algorithms with respect to their ability to predict release location and release mass of single/multiple source(s) combined with the actual number of source term locations/masses for each individual case, we decided to use this simple metric to capture high-level capabilities of individual algorithms. With respect to source location, this allowed us to compare trends among algorithms instead of trying to define a "weighted" combined release location/mass metric capable of penalizing individual algorithms based on the number and masses of predicted sources.



The distance between the predicted and the observed location for an individual source can, of course, be larger or smaller than the "miss distance metric" value that corresponds to an average difference when more than one location is involved in the release.

Figure 2. Example of the Distance Metric Computation and Mass Calculation used to Compare Algorithm Performance for each Individual Case

Comparison of Algorithms Based on Averaged Miss Distance

Of the 104 cases distributed to STE algorithm developers, the majority of the cases (80) were for single-source and double-source releases – 40 cases in each group. Hence, we focus our initial analysis on single- and double-source cases. Table 4 lists the composition of predicted single- and double-source cases that were provided by STE algorithm developers.

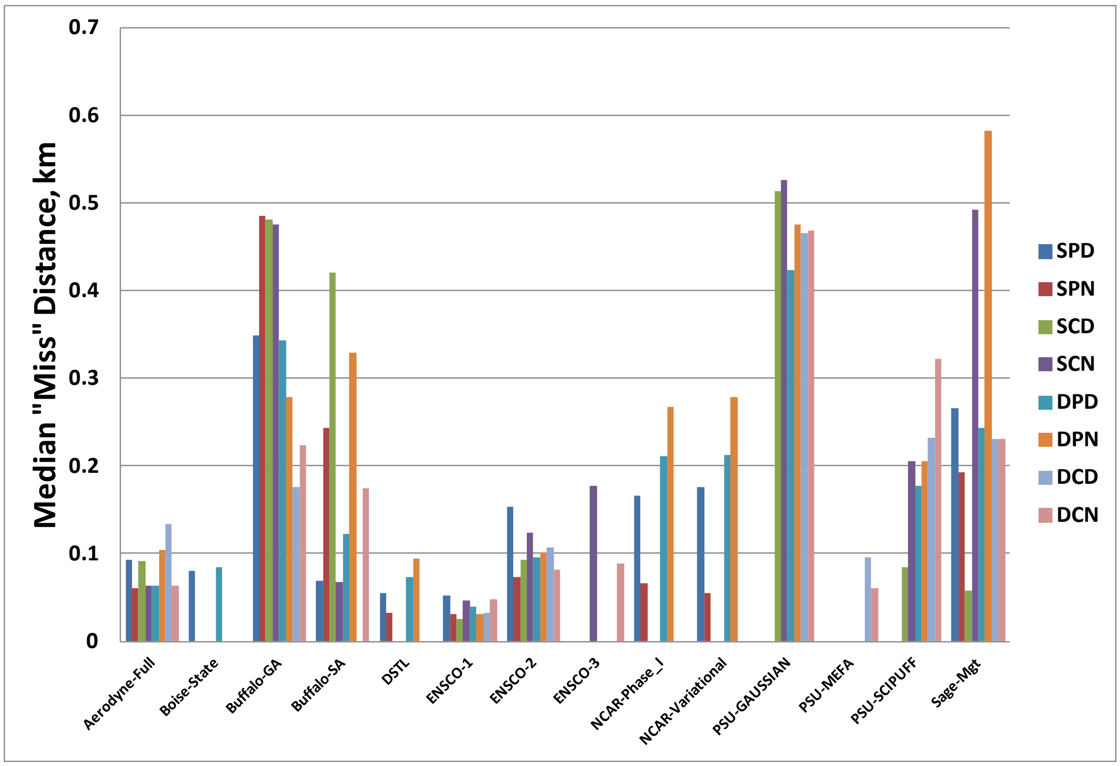
Table 4. Composition of Single and Double Source Predicted Cases Provided by STE Developers

Single							
Organization	Total	Cont	Puff	Puff/ Day	Puff/ Night	Count/ Day	Count/ Night
Aerodyne	40	20	20	10	10	10	10
Boise-State	13	5	8	5	3	3	2
Buffalo/GA	40	20	20	10	10	10	10
Buffalo/SA	26	11	15	8	7	6	5
DSTL	12	1	11	6	5	0	1
ENSCO/Set 1	39	19	20	10	10	9	10
ENSCO/Set 2	40	20	20	10	10	10	10
NCAR/Variational	24	9	15	6	9	6	3
NCAR/Phase I	24	9	15	6	9	6	3
Sage-Mgt	19	8	11	5	6	5	3
PSU/Gaussian	18	10	8	4	4	5	5
PSU/SCIPUFF	18	10	8	4	4	5	5
PSU/MEFA	13	6	7	3	4	3	3

Double							
Organization	Total	Cont	Puff	Puff/ Day	Puff/ Night	Cont/ Day	Cont/ Night
Aerodyne	40	20	20	10	10	10	10
Boise-State	13	5	8	5	3	4	1
Buffalo/GA	40	20	20	10	10	10	10
Buffalo/SA	26	14	12	6	6	4	10
DSTL	14	1	13	7	6	1	0
ENSCO/Set 1	39	20	19	9	10	10	10
ENSCO/Set 2	40	20	20	10	10	10	10
NCAR/Variational	19	3	16	8	8	2	1
NCAR/Phase I	19	3	16	8	8	2	1
Sage-Mgt	18	11	7	2	5	4	7
PSU/Gaussian	20	10	10	5	5	5	5
PSU/SCIPUFF	20	10	10	5	5	5	5
PSU/MEFA	16	10	6	4	2	5	5

Information is broken into several categories including release type, time of day, and number of sources. Red font values denote that a full set of predictions was provided; blue font values denote that the predictions were provided for at least 50 percent of the distributed cases.

Our expectation was that individual algorithm performance should be most affected by the number of sources, time of day of the release (e.g., daytime or nighttime), and type of the release (e.g., instantaneous or continuous).[4] That yields eight combinations (e.g., "Single Source"/Instantaneous/Daytime). In addition, to ensure adequate sampling of each individual grouping (with 80/8 = 10 cases per grouping), only algorithms that provided predictions for at least half of the distributed cases were included in each individual comparison. Figure 3 depicts algorithm performance broken down by these groupings in terms of the median miss distance, where the median is taken over all predicted cases in the subgroup.



Algorithms had to provide predictions for at least half of the cases to be included in each category listed in the legend. The first letter in the legend denotes number of sources: S – denotes a single source, D – denotes a double source; the second letter in the legend denotes the release type: P – denotes puff release, C – denotes continuous release; and the third letter denotes time of day of release: D – denotes a daytime release, N – denotes a nighttime release.

Figure 3. Median "Miss" Distance for Individual STE Algorithms

Differences in performance among algorithms are generally larger than differences in performance among release conditions within the set of predictions of an individual algorithm. Moreover, it is difficult to discern similar trends among different algorithms.

[4] Some of the algorithms were designed to function with subsets of the cases that were distributed (e.g., instantaneous and/or single sources.). In fact, several algorithm developers prescreened the available cases to try to select cases that corresponded to the design of their algorithm, but other developers with similarly limited algorithm designs decided to apply their algorithm to all cases.

Appendix E provides additional plots of miss distance for single and double releases grouped in various ways.

Linear Regression Analysis Results

We used stepwise and backward linear regression to examine which of the underlying factors – such as diurnal condition, the number of release sources, the type of release, and several other independent variables – had the greatest effect on the estimation of the mass ratio (the ratio of predicted to actual mass) or the miss distance.

Backward regression begins with all independent variables in the regression equation, and then proceeds to eliminate those for which the associated sum of squares is insignificant. In contrast, stepwise regression only allows independent variables into the regression equation if their associated sum of squares is significant and eliminates previously admitted variables if their effect is substantially diminished by other variables in the equation. Thus, roughly speaking, stepwise regression tests each independent variable to determine whether it should enter the regression equation, and again, if it should remain in the equation after others are admitted. Backward regression initially treats all variables as belonging to the equation, then eliminates those whose contribution is substandard [4, 5].

In this section we summarize results obtained using linear regression. Further details of the regression analysis are provided in Appendix F.

We chose the following independent regression variables:

1. "Diurnal," defined as either night or day release time
2. "Met Num," defined as either "close-in" met corresponding to meteorology obtained at the center of the sensor grid or "operational" met, which corresponded to using data from meteorological stations approximately 1-2 km away
3. "Sources," denoting the number of sources used in the definition of a case (single, double, triple, quad)
4. "Sensors," denoting the number of simulated sensors used in the definition of a case (4 or 16)
5. "Puff/Real," defined as "-1" if case is constructed from a continuous trial; "0" if case is constructed using single realization of a puff trial; and "1" if case is constructed using multiple realizations of a puff trial.[5] The "Puff/Real" inde-

[5] FFT 07 individual puff trials involved multiple (up to 10) realizations. These puffs were released by firing air cannons every few minutes resulting in "trains of puffs" periodically traversing the digiPID grid. Hence, for puff releases, some distributed cases included a single realization of the puff(s) and some included multiple (up to 10) realizations. The main idea of creating two types of releases based on puff trials was to exercise the STE algorithm's ability to temporally distinguish between single/multiple releases. Since this ability is only applicable to some STE algorithms, this potential analysis venue was left to individual STE developers.

pendent variable is expected to succinctly represent two distinct parameters that could affect quality of STE predictions: continuous versus instantaneous/puff releases and single versus multiple releases from the same location.

The list of dependent regression variables includes: “Mean,” defined as the distance between average predicted and average observed source term locations (as described earlier) for the individual case as shown in Figure 2, and “Mass Ratio,” defined as the ratio of predicted to observed total mass of the material used to define a particular case.

Results for stepwise and backward regressions are summarized in Tables 5 and 6 respectively. Each table is divided into two sections, one for each dependent variable. For each individual set of STE predictions, independent variables highlighted by regression as significant are marked by “x” and color coded to simplify viewing results.

With respect to predicting miss distance between predicted and observed STE location, the regression analysis indicates:

- The “Diurnal” (Day/Night) regression variable is not a significant variable in both backward and stepwise regressions.
- The “Met Num” regression variable representing “Close-In” versus “Operational” met options is not significant in both backward and stepwise regressions for almost all algorithms. The only exceptions are those submitted by ENSCO.
- The “Sources” regression variable representing number of sources used in the definition of a case is a significant predictor of algorithm performance for six algorithms. Sources are hence highlighted for six algorithms by stepwise regression and four algorithms by backward regression.
- The “Sensors” regression variable representing number of sensors (4 versus 16) used in the definition of the case is a significant predictor of algorithm performance for only three algorithms. This indicates that most STE algorithms do not benefit from being provided with data from a larger number of sensors.
- The “Puff Real” regression variable is a significant predictor of algorithm performance for two algorithms when using backward regression, and for one algorithm when using stepwise regression.

With respect to the “mass ratio” dependent variable, regression analysis indicates:

- The “Diurnal” (Day/Night), “Met Num” (Close-In/Operational MET), and “Sensors” (4 versus 16) regression variables are not significant variables for most algorithms for both backward and stepwise regression.
- The “Sources” independent regression variable representing the number of sources used in the definition of a case is a significant predictor of algorithm performance for seven algorithms.
- The “Puff Real” regression variable is a significant predictor of algorithm performance for seven algorithms.

Table 5. Table of Significant Factors for Backward Regression

Dependent Variable: Mass Ratio

Model	Diurnal	Met Num	Sources	Sensors	Puff Real
ENSCO 3			X		X
Buffalo SA	X	X	X		
DSTL			X		X
ENSCO 2			X	X	X
PSU Gaussian			X		X
PSU SCIPUFF			X		
Buffalo GA	X		X		X
ENSCO 1					X
Aerodyne				X	X
NCAR Phase I					
NCAR Variation					
SAGE Mgt August					
Boise State					
PSU MEFA					

Dependent Variable: Mean Distance

Model	Diurnal	Met Num	Sources	Sensors	Puff Real
ENSCO 3			X	X	
Buffalo SA					
DSTL			X		X
ENSCO 2		X		X	
PSU Gaussian			X		X
PSU SCIPUFF					
Buffalo GA					
ENSCO 1		X			
Aerodyne				X	
NCAR Phase I			X		
NCAR Variation			X		
SAGE Mgt August			X		
Boise State					
PSU MEFA					

(Column group heading for both tables: Independent Regression Variable)

Table 6. Table of Significant Factors for Stepwise Regression

Dependent Variable: Mass Ratio					
	Independent Regression Variable				
Model	**Diurnal**	**Met Num**	**Sources**	**Sensors**	**Puff Real**
ENSCO 3			X		X
Buffalo SA	X	X	X		
DSTL			X		X
ENSCO 2			X		X
PSU Gaussian					
PSU SCIPUFF			X		
Buffalo GA	X		X		
ENSCO 1					X
Aerodyne				X	X
NCAR Phase I					
NCAR Variation					
SAGE Mgt August					
Boise State					
PSU MEFA					

Dependent Variable: Mean Distance					
	Independent Regression Variable				
Model	**Diurnal**	**Met Num**	**Sources**	**Sensors**	**Puff Real**
ENSCO 3			X		
Buffalo SA					
DSTL					X
ENSCO 2					
PSU Gaussian					
PSU SCIPUFF					
Buffalo GA					
ENSCO 1		X			
Aerodyne				X	
NCAR Phase I			X		
NCAR Variation			X		
SAGE Mgt August			X		
Boise State					
PSU MEFA					

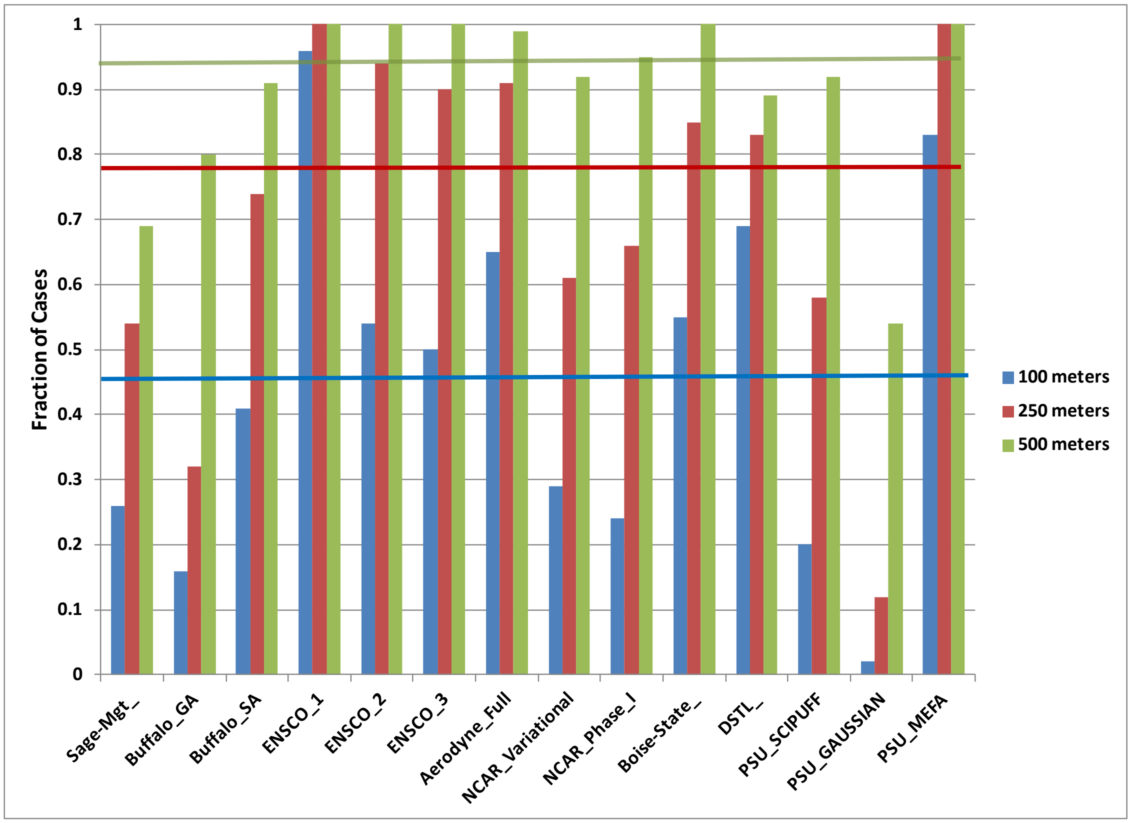
Early in this study, we conducted analyses of variance (ANOVA) of both the mass estimation and miss distance predictions with the intent of gaining insight into which of the many inputs to the various models had a significant effect on their outcomes. The results of the ANOVA indicated that, in certain cases, two-way interactions between factors (independent variables) were potentially significant. With these results as motivation, we reformulated the regression equations used earlier to include second-order terms, such as the product of the number of sensors and the number of sources. In certain cases, this required coding categorical variables, such as diurnal conditions, as scalar quantities (e.g., assigning the value 1 to daytime and -1 to nighttime). Thus, instead of attempting to "fit" outcomes to linear functions of several variables, we attempted to model outcomes as second-order polynomials in several variables. We then proceeded with stepwise regression and recorded the resulting adjusted R^2. Upon further examination of the results, we concluded that they were entirely consistent with linear regression results presented above. Appendix G depicts resulting tables.

We would like to caution that regression analysis results should serve as a guide for further investigation of which algorithm/variable combinations significantly influence predictive performance. For instance, regression analysis does not tell if the algorithm performed as expected with respect to a given variable.

Comparison of Selected Global Algorithm Performance Metrics

In addition to using the linear regression methodology to discern trends among different sets of STE predictions, we devised metrics to capture some aspects of global algorithm performance. As discussed earlier, for each individual case predicted by an STE algorithm, two measures were calculated: the distance between the average predicted and the average observed location of the source(s), which we will refer to as "miss distance"; and the ratio of total predicted mass to total released mass from all sources, which we will refer to as the "mass ratio."

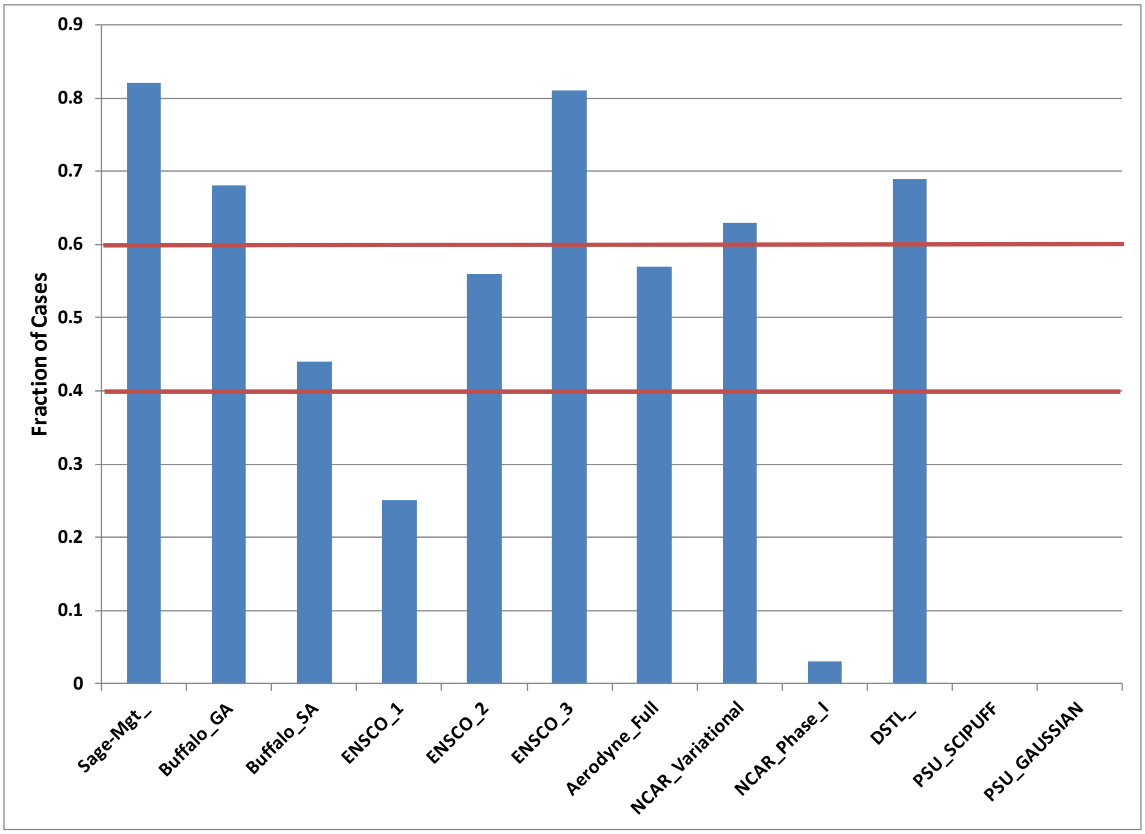
To compare STE algorithm performance using the miss distance metric, we selected three levels of interest and then calculated the fraction of cases in which miss distance is less than the level of interest. These levels of interest include: 100 meters (i.e., the miss distance is in the tens of meters), 250 meters (i.e., the miss distance is less than half the size of the sensor domain), and 500 meters (i.e., the miss distance is less than the approximate size of the sensor domain). We note that even when a particular miss distance is less than some number *d*, it is quite possible that the individual distances between actual and predicted locations of the sources is greater or less than *d*, as demonstrated in Figure 2. Figure 4 shows the results for these calculations at the three levels of interest. For each set of STE predictions, the grouped colored bars denote the fraction of predictions that are less than the particular level of interest. With respect to predicting miss distance, we observe the following: (1) when the miss distance is less than 100 meters, a wide spread is seen in algorithm performance; and (2) most algorithms seem to be capable of having more than 90 percent of their predictions have miss distances less than 500 meters (approximately the size of the tracer measurement grid of FFT 07).



Individual algorithm bars are color coded according to the legend. Thick colored lines correspond to the medians of fractions for all algorithms and at the various thresholds: 0.46 (blue line) for the fraction of miss distances less than 100 meters, 0.79 (brown line) for the fraction of miss distances less than 250 meters, and 0.94 (green line) for the fraction of miss distances less than 500 meters. Therefore, these lines separate algorithms into the better and worse performing halves, as measured by the given metric calculated over all cases for each algorithm.

Figure 4. Algorithm Inter-Comparison Using Averaged Miss Distance Fraction of Cases below 100, 200, and 500 Meters

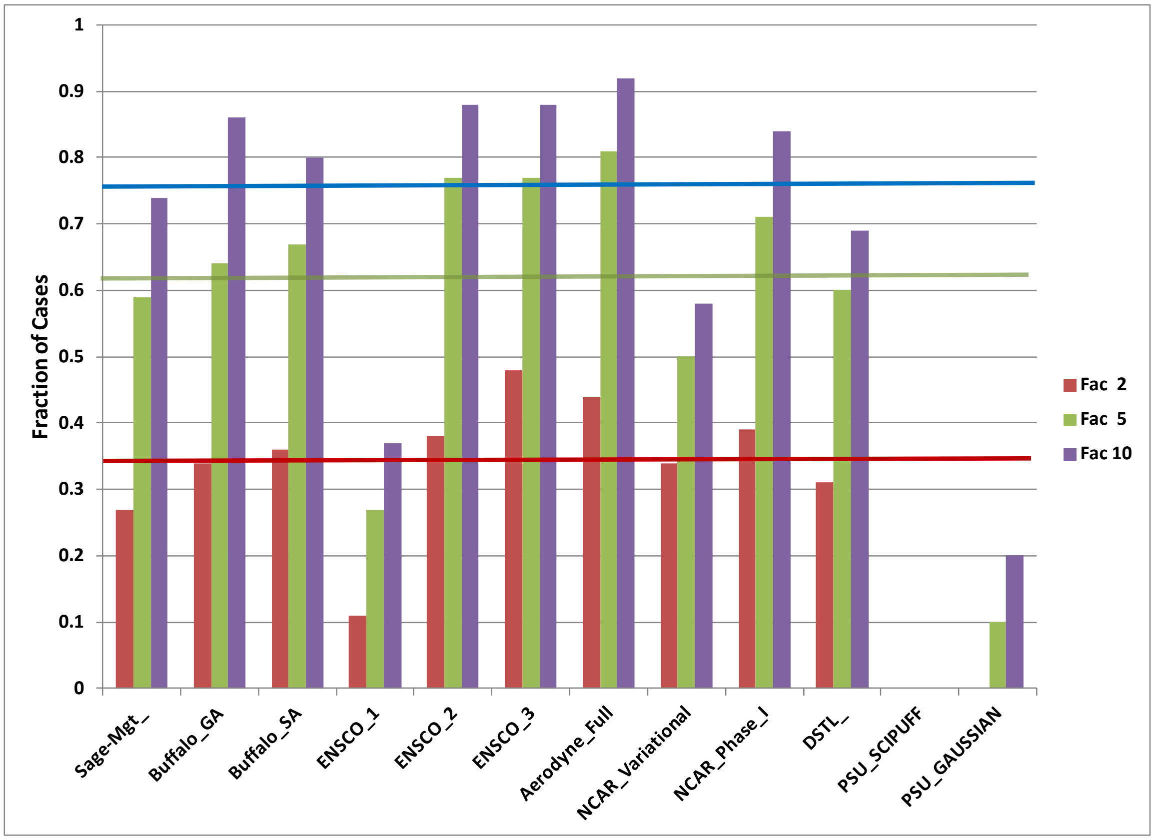
We examined the mass ratio metric for two types of statistics: (1) whether a particular algorithm has a tendency to over- or under-predict the total mass released from all sources, and (2) for any given set of predictions, what is the fraction of the cases when the predicted and observed masses are within a factor of 2, 5, or 10 of each other. For each set of the 12 predictions that provided enough information to calculate the total predicted mass, Figure 5 shows the fraction of cases that were over-predicted.



Thick brown lines (at 0.4 and 0.6) denote limits that are used to distinguish different predictive behavior: a fraction below 0.4 implies an algorithm tendency to under-predict, a fraction in the range of 0.4 and 0.6 implies about an equal number of under- and over-predicted cases, and a fraction above 0.6 implies an algorithm tendency to over-predict.

Figure 5. Total Mass Over-Prediction Fraction for the 12 STE Algorithms that Provided Enough Information to Calculate Total Predicted Release Mass from All Sources

For each set of predictions, Figure 6 shows the fractions of cases in which the total observed and predicted masses are within factors of 2, 5, and 10 of each other. With respect to the total predicted-to-observed mass ratio metric, we observe the following: (1) wide variations appear in terms of algorithm performance with respect to over- or under-predicting masses of the releases, with some algorithms exhibiting a large number of cases significantly over- or under-predicted; (2) with the exception of three algorithms, the fraction of cases in which the predicted total source mass fell within factors of 2, 5, or 10 of the actual total source mass varies from 0.27 to 0.48, 0.59 to 0.81, and 0.69 to 0.92, respectively.



Thick colored lines correspond to the medians of fractions for all algorithms and at the various thresholds: 0.34 (brown line) for factor of 2, 0.62 (green line) for factor of 5, and 0.77 (blue line) for factor of 10.

Figure 6. Algorithm Inter-Comparison Using Observed and Predicted Mass Fractions within Factors of 2, 5, and 10 of Each Other

We caution that these results capture global algorithm performance without any attempt to ensure that the compared predictions are compatible with each other. For instance, these results do not take into account that some algorithms provided only partial predictions (i.e., not a complete set of predictions for all cases). Some of the algorithm developers preferentially selected sets of predictions to submit (e.g., "Puff only" or "16 sensors only" predictions).

A. Discussion

With respect to our miss distance metric, all algorithms were able to predict "averaged" source term locations to within 500 meters (i.e., a size comparable to the size of the tracer measurement grid of the FFT 07 experiment); a wide variation in the quality of the algorithm predictions was seen when the miss distance was on the order of tens of meters (i.e., less than 100 meters). Few algorithms are able to consistently predict the source of a release with an accuracy of more than a few hundred meters. We note that the FFT 07 sensor grid was less than 500 meters across and that the release sources were less than 100 meters away from the leading edge of the sensor grid.

With respect to predicting the total release mass, a wide variation appears in algorithm performance with respect to over- or under-predicting masses of the release, with some algorithms showing large fractions of cases that were under-predicted and some showing large fractions that were over-predicted. About half of the models were able to predict the total mass of the source to within a factor of 10 of the actual source mass for about three-quarters of the cases. When the prediction standard quality was raised to within a factor of 2, about half of the algorithms had this level of accuracy for less than one-third of the cases. Most of the STE algorithms that were evaluated cannot consistently predict the total mass to within a factor of 5 of the actual mass release. We would like to caution that these results are an attempt to capture global algorithm performance without any attempt to ensure that the compared predictions are compatible with each other. For instance, as noted earlier, these results do not take into account that some algorithms provided only partial predictions (i.e., not a complete set of predictions for all cases), and some of the algorithm developers preferentially selected sets of predictions to submit.

Linear regression analysis indicated that the time of the release (night versus day), type of meteorology provided (detailed versus sparse "operational" meteorology), and number of simulated sensors (4 versus 16) *did not* lead to significant differences in prediction quality for most of the STE algorithms under evaluation. Some confirmation of algorithm insensitivity to variations of the input data could be discerned by careful examination of predictions for individual cases supplied by each individual algorithm, although it is more difficult to quantify trends among all algorithms by examining algorithm predictions of individual cases. At first glance, this result seems to be counterintuitive. For instance, one expects that quadrupling the number of sensors from 4 to 16, or using high-frequency close-in meteorology, should necessarily lead to better predictions. Also, the time of the release (e.g., daytime versus nighttime), in general, has a strong correlation with the atmospheric stability, which should significantly affect atmospheric dispersion. Thus, it is rather unexpected that STE algorithms are capable of predicting source term parameters with equal skill under stable and unstable atmospheric conditions. We speculate that the relatively small spatial scale of the FFT 07 digiPID sensor grid (approximately 450 by 450 meters) and the proximity of release locations to each other and to the upwind leading edge of the sensor grid are responsible for this. For instance, for most single-source releases, the cross-wind extent of the plume does not cover more than few neighboring digiPIDs, and no significant spatial variation occurs in the plume over the sensor grid as the downwind distance from the release location increases. Thus, changing the number of simulated sensors from 4 to 16 might not provide enough additional information for the STE algorithms.

Linear regression analysis also indicated that the number of sources and type of release [continuous release versus single realization of instantaneous puff(s) versus multiple realizations of instantaneous puff(s)] are significant variables in terms of predicting algorithm performance for the majority of participating algorithms. We note that regression analysis itself (as used here) does not quantify the quality of the algorithms' ability to predict source term parameters – it only indicates which release factors have an effect on the quality of the STE predictions.

Our most significant observations and recommendations from these investigations are described below:

- **Source term estimation, as envisioned for chemical and biological weapon attacks, remains a challenge.** An initial look at state-of-the-art STE algorithms participating in this exercise revealed potential shortcomings with respect to estimating the spatial location and mass of the release. Although most STE algorithms seemed capable of estimating the location of the release on a scale comparable to the limited size of the sensor grid used in FFT 07, and noting that the releases were very close to the upwind edge of the sensor grid, questions remain as to how well these algorithms would perform using operationally relevant scenarios including sensors that are spaced farther apart from each other and the release location.[6]
- **The FFT 07 field trials appear to have limited applicability to practical validation of STE algorithms.** FFT 07 is the most comprehensive field trial conducted to provide information to further the development and assessment of STE algorithms – certainly a valuable and necessary source of measurements and observations for this goal of improving the state of the art.[7] However, the relatively small size of the sensor grid and the closeness of the release locations to the upwind leading edge of the sensor grid, limit the usefulness of FFT 07 as the basis for any future validation of an STE algorithm for militarily relevant scenarios. Moreover, our analysis revealed that certain input variations for the STE algorithms (such as quadrupling the number of available sensors or providing detailed high-resolution meteorology near the center of the sensor grid) did not lead to expected discernible improvements in the quality of the STE predictions. This suggests that the small scale of FFT 07 – a few hundred meters – limited its usefulness for evaluations of even fundamental STE algorithm performance at larger (and for many applications, more realistic)

[6] One could conceive of an STE algorithm that places the source at the location of the first sensor that detects the release. This type of algorithm would be consistent with placing an Allied Tactical Publication-45 (ATP-45) warning triangle at the sensor that registers first detection. Given the limitations associated with the FFT 07 field experiment (especially the scale), such an algorithm would perform quite comparably to the more complex STE algorithms that were investigated.

[7] The provided FFT 07 data were quite valuable to algorithm developers, especially in terms of refining their expectations. For instance, several prototype algorithms did not expect that (1) some sensors have "noise" floors (i.e., they register some signals even when no tracer gas was present), and (2) different sensors have differing levels of noise. That necessitated some developers to implement new threshold algorithms before supplying the provided data to their algorithms.

scales where atmospheric stability, the quality of meteorological inputs, and the amount of available sensor (i.e., "detector") information can reasonably be hypothesized to influence STE algorithm performance.

- **A relatively high-fidelity, virtual, simulated environment could be useful for future assessments, and even independent validation activities, of STE algorithms.** Of course, this recommendation rests on premise that a relatively large-scale, realistic field trial is unaffordable (and possibly not executable in any case). As computational power becomes more available and relatively cheap, the potential exists to use computer modeling tools to supplement field testing of system components. The use of such tools holds the promise of increasing the efficiency of the field tests that are conducted, aiding the evaluation of results obtained from such tests, and reducing costs. We recommend that simulated environments such as the National Center for Atmospheric Research (NCAR) Virtual THreat Response Emulation and Analysis Testbed (VTHREAT) modeling system should be considered and take central stage to supplement and extend field trial data. Furthermore, if future assessment and validation efforts of STE modules will largely (and probably appropriately) rely on simulated environments, future laboratory measurements or field trial designs and observations must take this into account. That is, we recommend a holistic approach to designing the strategy by which simulated environments and field trials (or laboratory tests) are used to further the assessment and validation of STE modules. Such an approach should ensure future activities are complementary and should especially seek synergistic activities (e.g., field trial or laboratory observations that support increased confidence in aspects of the virtual environment that are critical to its use when applied to STE algorithm assessment).

References

1. Donald P., Jr., Storwold, *Detailed Test Plan for the Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07)*, Meteorology Division, West Desert Test Center, U.S. Army Dugway Proving Ground, September 2007.
2. Platt, N., S. Warner, and S.M. Nunes, *Plan for initial comparative investigation of source term estimation algorithms using FUSION field trial 2007 (FFT 07)*. IDA Document D-3488, 2008.
3. Platt, N., S. Warner, and S.M. Nunes, *Evaluation plan for comparative investigation of source term estimation algorithms using FUSION field trial 2007 data*, Croatian Meteorological Journal, Proceedings of the 12th International Conference on Harmonization within Atmospheric Dispersion Modeling for Regulatory Purposes, Part 1: Oral Presentations, Vol. 43, p. 224-229, 2008.
4. Seber, G., *Linear Regression Analysis*, Wiley, 1977.
5. Draper, N. and H. Smith, *Applied Regression Analysis*, Wiley, 1966.
6. Bieberbach, G., P.E. Bieringer, A. Wyszogrodzki, J.Weil, R. Cabell, J. Hurst, and J. Hannan, *Virtual chemical and biological (CB) agent data set generation to support the evaluation of CB contamination avoidance systems*, The Fifth Symposium on Computational Wind Engineering (CWE 2010), Chapel Hill, North Carolina, 2010.

(This page is intentionally blank.)

Appendix A
Abbreviation

AIMS	Aerodyne Inverse Modeling System
AMS	American Meteorological Society
ANOVA	analysis of variance
ARI	Aerodyne Research, Inc.
ATD	atmospheric transport and dispersion
ATP-45	Allied Tactical Publication-45
BIC	Bayesian Information Criterion
CB	chemical and biological
CBD	Chemical and Biological Defense
CBRN	chemical, biological, radiological, and nuclear
CONOPS	Concept of operations
digiPID	Digital Photoionization Detector
DoD	Department of Defense
DPG	Dugway Proving Ground
DSTL	Defense Science and Technology Laboratory (United Kingdom)
DTRA	Defense Threat Reduction Agency
FBT	Forward-Backward Trajectory
FFT	Fusion Field Trial
FFT 07	Fusing Sensor Information from Observing Networks Field Trial 2007
4DVar	Four dimensional variational
FUSION	Fusing Sensor Information from Observing Networks
GA	genetic algorithm
GA-Var	Genetic Algorithm variational
GMU	George Mason University
H-LEPM	Hybrid-Langrangian-Eurelrian Model
IDA	Institute for Defense Analyses
JEM	Joint Effects Model
JPO	Joint Program Office
JSTO	Joint Science and Technology Office
JSTO-CBD	JSTO for Chemical and Biological Defense

kg	kilograms
km	kilometers
LR	Linear Regression
LR-sub	Linear Regression-subset
m	meters
MCBDF	Monte Carlo Bayesian Data Fusion
MCMC	Markov Chain Monte Carlo
MEFA	Multiple Enitity Field Approximation
MO	Monin-Obukhov
MOE	Measure of effectiveness
NCAR	National Center for Atmospheric Research
NSWCDD	Naval Surface Warfare Center Dahlgren Division
NYC	New York City
PWIDS	Portable Weather Information and Display System
S&T	Science and Technology
SA	simulated annealing
SCIPUFF	Second-Order Closure Integrated Puff
SDF	Sensor Data Fusion
SERT	Stochastic Event Reconstruction Tool
STE	Source Term Estimation
TP9	Technical Panel 9 for Hazard Assessment
T&D	transport and dispersion
TP	Technical Panel
TTCP	The Technical Cooperation Program
UK	United Kindgdom
UDP	Urban Dispersion Program
UVIC	ultraviolet ion collector
VTHREAT Testbed	Virtual THreat Response Emulation and Analysis
V&V	Verification and Validation

Appendix B
Sequence of Events

The following is a brief summary of the sequence of events that took place before this document was written and that are related to the evaluations described in this document:

1. IDA and DPG held a meeting at DPG in late October 2007 to discuss plans and to structure the proposed exercise as instructed by our DTRA sponsor. DPG was responsible for running the FFT 07 field trial and subsequent data management and distribution. IDA agreed to create an evaluation plan.
2. A draft version of the evaluation plan was briefed to the FFT 07 science team in December 2007.
3. A draft version of the evaluation plan was distributed to potential STE participants and the FFT 07 science team in January 2008.
4. The draft evaluation plan was briefed at the annual TP9 meeting in February 2008 with feedback requested from the STE developers.
5. A final version of the evaluation plan, incorporating the changes agreed to among STE algorithm developers, the DTRA sponsor, and members of FFT 07 science team, was distributed in May 2008.
6. Processed DigiPID data were received at IDA in June 2008.
7. The FFT 07 science team held a side-bar meeting during the George Mason University (GMU) ATD conference in July 2008.
8. Cases of simulated sensor data were made available to the STE developers in September 2008.
9. IDA initiated and attended a series of one-on-one meetings with interested STE algorithm developers during October and November 2008.
10. Preliminary predictions were received at IDA in December 2008.
11. Preliminary results of IDA analyses were submitted in December 2008 and presented at the annual TP9 meeting in February 2009. With sponsor's

concurrence and STE developers' request, the deadline to submit a final set of predictions was extended until the end of August 2009.

12. IDA briefed results of the preliminary analysis of STE algorithm performance, including prediction updates received since December 2008, at the GMU ATD conference in July 2009.
13. At the end of August 2009, the exercise was officially closed with respect to submitting updates to predictions.
14. Developer feedback package was prepared and distributed to STE algorithm developers in September 2009.
15. IDA briefed results of the analysis at the annual TP9 meeting in September 2009 and at the annual American Meteorological Society (AMS) meeting in January 2010.

Appendix C
Brief Description of Source Term Estimation Algorithms

This appendix is devoted to a brief description of STE algorithms that provided predictions for this investigation. Eight organizations provided 14 sets of full and partial predictions. Some organizations provided multiple sets corresponding to different algorithms they were developing, or, in one case, to an increased size of the spatial search box used within their algorithm.

All STE algorithm descriptions were provided by STE algorithm developers with minor editing done by IDA. We thank the developers who responded to our request to provide this information. The rest of this appendix is organized into subsections corresponding to the individual organizations that participated in the exercise.

A. Predictions Provided by Aerodyne Research, Inc. (denoted "Aerodyne")

Aerodyne Research, Inc. (ARI) developed an algorithm for source term estimation, called *AIMS* ("Aerodyne Inverse Modeling System"). In general terms, AIMS applies a variational approach for source estimation: a cost function is defined that quantifies the mismatch between all observations and the corresponding model predictions resulting from a given set of trial source parameters; then, the optimal set of source parameters is identified as the values for which the cost function is minimized (see Equation 1).

$$\begin{aligned} &Cost(\beta) = \left\| Data - Model(\beta) \right\| \\ &\beta^* = \arg\min_{\beta} Cost(\beta) \end{aligned} \tag{1}$$

where β is the set of unknown source parameters; and β^* is the value of β that yields forward model predictions that are most consistent with the data.

Indeed, in the theoretical limit of ideal data and models, the global minimum of this cost function exists at the set of parameters that is most likely responsible for the observational data. The two main challenges of variational approaches in practice involve successfully locating the (global) minimum of the cost function and dealing with non-ideal data and models. The former challenge demands careful definition of the cost function and the use of a robust minimization algorithm. The latter requires awareness of (and accounting for) artificial offsets in the location of the minimum, due to non-ideal data and models. Approaches for addressing these issues are detailed in upcoming ARI papers.

AIMS takes as input all available observational data and optionally any prior knowledge of the source parameters. The output is the set of source parameters that best describes the observations, including number of sources, emission rates, locations, and start and end times. AIMS is also designed to include an *a posteriori* assessment of its solution quality, providing useful feedback on how much confidence to put in a particular solution and in what ways the solution quality might be improved.

A novel feature in AIMS is the ability to integrate multiple observation types in order to maximize information content for source estimation. This capability has been demonstrated for datasets from stationary and mobile sensors.

References

1. S.E. Albo, O.O. Oluwole, R.C. Miake-Lye, "The Aerodyne Inverse Modeling System (AIMS): Source estimation applied to the FFT 07 experiment and to simulated mobile sensor data," in preparation, *Atmospheric Environment*, 45, p. 6085-6092, 2011.
2. O.O. Oluwole, S.E. Albo, R.C. Miake-Lye, *Source estimation using SCIPUFF Tangent-Linear or Adjoint*, CBD Physical Science and Technology conference proceedings, 2008.

B. Predictions Provided by Boise State University (denoted "Boise State")

The Stochastic Event Reconstruction Tool (SERT) adopts a probabilistic approach that delivers results with uncertainty quantification. The probabilistic approach is based on Bayesian inference with Markov Chain Monte Carlo (MCMC) sampling (Senocak et al., 2008). SERT is computationally fast and runs in minutes on a laptop. *The current version of SERT is designed to address continuous releases from a single source using a stochastically enhanced Gaussian plume model*. However, it was applied "as is" to puff and multiple source releases during the FFT 07 Phase 1 blind evaluation study. Ideally, multiple source dispersion models and puff models for instantaneous releases should be implemented in SERT. Novel features of the SERT code can be listed as follows:

- Given the sensor data, empirical parameters in the dispersion model are estimated stochastically using the Bayesian inference engine. The practice improves results tremendously and optimizes the dispersion model for each specific problem at hand.
- SERT directly incorporates the sensitivity of chemical and biological (CB) sensors/collectors. Trace amounts of CB agents may not be detected by a sensor because of its detection sensitivity governed by a concentration threshold.

Therefore, SERT does not ignore zero-hit sensors. It incorporates the information into the probability model of the Bayesian inference engine by attaching a probability to zero sensors.

- SERT solves the inverse problem with as many as nine distinct parameters (e.g., source location, strength, wind direction, wind speed, and turbulence diffusion parameters) simultaneously. Results are always delivered with uncertainty quantification, which is an inherent feature of the Bayesian inference method.
- SERT does not have problem-specific tunable parameters. SERT estimates all the parameters in a principled way using prior probability distributions.
- SERT code is written in JAVA. Forward plume models for different dispersion scenarios can easily be added thanks to the object-oriented software design.

References

1. Senocak, I., N.W. Hengartner, M. Short, B. Daniel, “Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference,” *Atmospheric Environment*, Vol. 42, 7718-7727, 2008.

C. Predictions Provided by University AT Buffalo (denoted “Buffalo/GA” and “Buffalo/SA”)

Data collected during FFT 07 were used for developing STE algorithms for atmospheric chemical dispersion. Heuristic approaches such as simulated annealing (SA) and genetic algorithms (GA) are used. The developed STE algorithms provide the best estimates of the source locations, source type (continuous/single puff/train of puffs), source strengths, number of sources, release start time, and end time.

Second-Order Closure Integrated Puff (SCIPUFF) is used as the predictive model for the atmospheric dispersion process. The source parameters are estimated by minimizing the cost function, which is the sum of the squared errors between model predictions and the given concentration sensor data at various sensor locations for all times. The STE algorithm is run in a post-processing mode, assuming a maximum of four sources. The actual number of sources is selected based on the Bayesian Information Criterion (BIC).

The given surface and profile wind sensor data are used to drive the predictive model. The sonic data and concentration data are reduced to 10-second data using backward averaging. However, turbulence calculations are not performed. Concentrations less than 0.001 kg/m^3 are neglected in the cost function evaluation. Some assumptions are made to reduce the search space during optimization. The maximum source strength is fixed at 10 kg for instantaneous and 1,000 L/min for continuous sources. In the case of multiple sources, all sources are assumed to be released at the same time and stopped at the same time. For a train-of-puff release, the time separation between successive puffs is assumed to be constant and is at least 1 minute long.

The release type is identified based on the plot of peak concentrations at various times. The top few concentration peaks and their neighborhood sensors are identified. If peak concentrations across this group of sensors stay above a certain concentration level for more than 2 minutes, then the release type is identified as continuous; if not, they are considered to be a puff release. For a puff release, the number of puffs and the time separation between successive puff releases can be identified approximately based on the number of peaks and the time separation between peaks. For a continuous release, the duration of release can be estimated approximately based on how long the peak concentration is above a certain level. The release time is assumed to be between the first measurement time in the given noisy concentration data and the time corresponding to the first concentration peak. Based on the wind variability, the bounds on possible source locations are estimated. The input files to run SCIPUFF are then prepared, and the model is started with an initial guess of source locations, strengths, and release start time. The minimization of the cost function is performed using the heuristic methods (SA/GA), assuming a maximum of four sources. The model with the lowest value of BIC is the preferred one.

For SA approach, 70 (of 104) cases are submitted for evaluation. The optimization methods do not have gradient information of the SCIPUFF model. Hence the time required to reach the global optimum is usually high, depending on tuning parameters: 20 minutes assuming single source for SA (and 2 hours to evaluate for up to four sources and select one).

D. Predictions Provided by Defense Science and Technology Laboratory, UK (Denoted "DSTL")

DSTL's Monte Carlo Bayesian Data Fusion (MCBDF) algorithm is a Bayesian posterior probability sampling algorithm that constantly updates its inference on release source terms conditional upon continuously arriving data.

A fixed-sized time window is maintained in which data are considered. Old data are discarded as time advances. This allows for real-time inference given sufficient computing power. In between the arrival of data, MCMC sampling is used to propose and possibly accept new hypothesized releases conditional upon the existing dataset. Dispersion code output for each proposed release is calculated and stored in an efficient manner for reuse. Upon the arrival of new data, each existing hypothesis has its weight multiplied by the likelihood of the data. The combination of parameters and weights encodes the posterior probability distribution from which inferences can be made. (The full details of the algorithm are given in the reference below.)

The parameter space used for the FFT 07 analysis described instantaneous point releases and had nine dimensions: two for the release location (release height was fixed), time of day of the release, mass, material (redundant in this analysis), northerly and easterly components of a spatially and temporally homogeneous horizontal wind vector, surface roughness length, and Monin-Obukhov (MO) length.

The prior distributions used for this analysis were uniform on location within a 2-km square centered on the sensors, uniform on time for 5 minutes before current time, exponential on mass with a mean of 100 kg, normal (variance 10 $m^2 \cdot s^{-2}$) on the wind components, and uniform on the log of the surface roughness and the reciprocal of the MO length.

Two likelihood models were used for the FFT 07 analysis. The concentration sensor model was a simple, normally distributed measurement error model. DSTL's Urban Dispersion Model was used to link the release parameter space to the concentration probability distribution at each measurement location and time. The unobserved concentration was integrated out. The wind measurement model used a bivariate normal component likelihood with a measurement covariance derived from the high-frequency variations.

References

1. P. Robins, V.E. Rapley, N. Green, "Real-time sequential inference of static parameters with expensive likelihood calculations," *JRSS Series C*, Vol. 58, Issue 5, pp. 641-622, December 2009.

E. Predictions Provided by ENSCO, Inc. (denoted "ENSCO 1," "ENSCO 2," and "ENSCO 3")

ENSCO offered two separate approaches to address the source term location and characterization challenges posed by the FFT 07 propylene release field experiment. The first approach ["ENSCO 2" hereafter referred to as Linear Regression (LR) and "ENSCO 3" hereafter referred to as Linear Regression-subset (LR-sub) datasets[1]] employed a linear regression methodology using releases from a grid of virtual sources to estimate source location. The second method ["ENSCO 1" hereafter referred to as Forward-Backward Trajectory (FBT) dataset] represents more of a holistic approach that integrates most components of available sensor and meteorological input data collected during FFT 07 with extensive subject matter expertise in atmospheric signal analysis. Neither

[1] "ENSCO 3" set (LR-sub) of predictions extended the limited spatial search box used in the "ENSCO 2" (LR) set of predictions and, due to time and budget constraints, was run for a subset of Phase I cases.

method need be tied to a particular transport and diffusion model. Depending on the preference of the user, any legitimate model could be used.

The LR approach uses a simple transport and dispersion model to generate emissions from the grid of virtual sources and correlates the predicted signals to the observed signals across the array of reporting digiPIDs. The linear regression model (Neter and Wasserman, 1974) had been applied previously in a long-range transport study (Masters, 1988).

For each virtual source, each release time, the regression model takes the form of:

$$\mathbf{Y} = b_0 + b_1\mathbf{X} + e \tag{2}$$

$\mathbf{Y}$ = Observed concentrations (all samplers, all collection times in range of release time)

$\mathbf{X}$ = Model predicted concentrations

b_0 = Regression intercept term (not used)

b_1 = Regression slope term, interpreted as the release rate for the source and time

e = Residual (error).

The result is a series of grids of correlations and slope terms across the virtual source grid at each possible release time. A set of thresholds is applied (e.g., correlation >0.7, slope term <1.0), which selects a subset of the space of the source and release time.

Highly correlated source grid locations are binned by release time to determine the most likely location for a source or sources. The higher the number of release times that correlate with a particular grid source, the more likely it is that the location is at or near a real source. After all information from all virtual sources is processed, a “weighted centroid” is calculated to identify an actual source location. The method appears to work well for single sources (burst or continuous) but currently will only identify the mean location of multiple sources, i.e., the centroid is likely to be near the center of a grouping of two or more sources. Additional work with clustering algorithms could facilitate the separation of source locations when more than one source is present.

The FBT method emphasizes the inclusion of only the most statistically significant points in the data stream. The algorithm defines a statistical noise threshold above which a measurement is considered to be a “plume.” By definition, such points, when connected to forward and/or backward trajectories, are much more likely to represent centerline, or near-centerline, hits that provide a very good first estimate of the azimuth of a source. This is particularly true if peaks at upwind and/or downwind sensors are highly correlated. The method constructs trajectories in time originating from as many digiPID locations as are represented by peak hits. Given there is at least minimal temporal variability of the wind field (even as little as 5-10 degrees), backward trajectories will intersect at or near a common point representing the location of a source. These analyses often readily reveal not only single sources, but the existence and location of multiple sources. Until the method can be fully automated using convergence routines tailored to this purpose, some minor semi-subjective nudging of trajectories may be necessary to best place trajectories that are not quite centerline hits. The direction of such adjustments is dictated by the nature of signals observed at upwind and/or downwind digiPIDs.

The value of the second approach is that it requires only a subset of all data and uses only those points that intrinsically provide the most complete information. Success is not dependent upon brute force calculation, but results can be improved by using a transport and dispersion model best suited for the synoptic situation and scale of transport. Principally, this method was conceived to offer the best opportunity to identify source locations with the premise that no transport model, regardless of sophistication, is of much use if the source(s) is/are determined to be in the wrong location(s).

References

1. Masters, S., *Source identification using meteorological and statistical modeling*, Preprints 10th Joint Conf. on the Applications of Air Pollution Meteorology with A&WMA, 11-16 January 1998, Phoenix, AZ, 1998.

2. Neter, J. and W. Wasserman, *Applied Linear Statistical Models*, Homewood, IL, Irwin, 1974.

F. **Predictions Provided by NCAR** (denoted "NCAR Variational" and "NCAR Phase I")

NCAR, under DTRA JSTO sponsorship, is one of a group of research organizations developing a CB sensor data fusion (SDF) algorithm package. This algorithm is required to estimate source term characteristics and provide a refined downwind hazard prediction, based on available CB and meteorological sensor measurements.

This algorithm uses variational data assimilation techniques in conjunction with a Gaussian puff dispersion model and an inverse plume modeling method to better characterize the source parameters and improve the accuracy of the subsequent plume dispersion solution. It leverages the relative strengths of the both the inverse plume modeling and variational approaches to address the atmospheric CB release source estimation problem. The major components of this algorithm are depicted in Figure C-1. The algorithm consists of a pre-processing step, a technique for making a first guess for the source type – SCIPUFF, its corresponding STE model – a simplified Hybrid-Lagrangian-Eulerian Plume Model (H-LEPM), its numerical adjoint, and the software infrastructure necessary to link them. SCIPUFF and its STE model are used to calculate a "first guess" source estimate based on the available CB and meteorological observations and source type estimation, denoted by "NCAR Phase I." The H-LEPM and corresponding adjoint are then used to iteratively refine the SCIPUFF-based STE estimate using variational data assimilation techniques. The entire process from beginning to end is completely automated and requires no human intervention. The algorithm is designed to be run on a laptop computer and provide a set of source parameters from seconds to several minutes after observations are provided to the algorithm. The technique is suitable for any atmospheric transport and dispersion (T&D) application where concentration observation and meteorological data are available and one or more of the release source parameters are not known. This methodology is particularly applicable for emergency response applications involving the dispersion of hazardous materials where a T&D solution is required as soon as possible following the collection of observations.

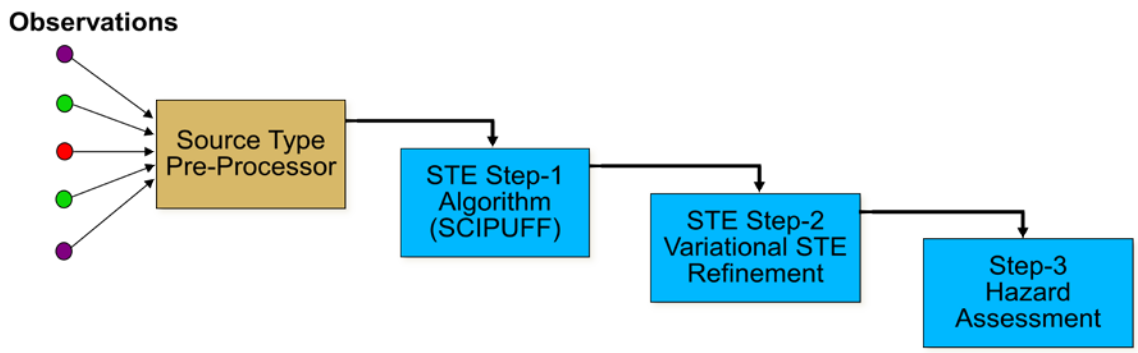


Figure C-1. The NCAR/Sage Management Variational STE and Hazard Refinement Algorithm Data Flow Design

References

1. P.E. Bieringer, I. Sykes, F. Vandenberghe, J. Hurst, J. Weil, G. Bieberbach, S. Parker and R. Cabell, *Automated Source Parameter Estimation for Atmospheric, Transport and Dispersion Applications*, Proceedings of the 13th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Paris, France, 2010.

G. Predictions Provided by Sage Management (denoted "Sage Mgt")

Sage Management's STE algorithm uses an adjoint SCIPUFF methodology for estimation of source term parameters. The adjoint release from each sensor measurement provides an estimate of the actual release mass at all prior upwind locations. The search methodology finds the location, in time and space, where optimal consistency exists among the different release mass estimates from all sensors, including null observations.

Several model improvements were implemented during the exercise, including completion of the treatment of a probabilistic estimate for continuous sources and an adjustment to the weighting function for null sensors.

Meteorological and sensor data were averaged with fixed averaging times, determined by trial and error to be appropriate for the instrumentation and travel times in question. Both instantaneous and continuous source searches were performed, and the optimum estimate was determined from the best forward predictions using an objective error measure.

We note that the adjoint SCIPUFF methodology is restricted to single source searches, so the multiple release cases, which form the majority of the FFT 07 cases, are strictly beyond the capability of SCIPUFF. Subjective examination of some of the test cases suggests that multiple sequential releases can sometimes be reasonably represented as a single continuous release, but multiple locations produce inconsistent results and generally force the locations estimate too far upwind.

References

1. Sykes, R.I., *Source Estimation using Reverse Transport*, presentation at 74th MORSS, US Air Force Academy, Colorado Springs, CO, 2005.
2. Fry, R., R.I. Sykes, and R. Kolbe, *Chemical/Biological Source Characterization*, presentation at Science and Technology for CBIS, Albuquerque, NM, 2005.
3. Sykes, R.I., *Source Estimation using Sensor Data and Reverse Transport*, presentation at Science and Technology for CBIS, Austin, TX, 2007.

H. Predictions Provided by Penn State University (denoted "PSU GAUSSIAN," "PSU SCIPUFF," and "PSU MEFA")

The Penn State assimilation team uses two different primary approaches to back-calculating source and meteorological data given field-monitored concentrations: GA-Var and a Multiple Entity Field Approximation (MEFA). Although we emphasize back-calculation of source strength, source location, release height, and time of release, experience has shown that the solutions are highly sensitive to errors in meteorological variables, so we have also back-calculated wind speed, wind direction, depth of the boundary layer, and stability variables.

Genetic algorithm-variational (GA-Var) uses a real-valued GA in a similar way to the variational approaches to data assimilation. It avoids the backward integration step of traditional four-dimensional variational (4DVar) techniques by directly optimizing the unknown variables using forward integration and solution evolution. The method relies on the GA operations of selection, mating, and mutation to provide a robust approach that is capable of finding global solutions to difficult optimization problems. We have developed GA-Var over a period of years, tested it for back-calculating all source and meteorological variables listed above, as well as using it for sensitivity studies of how much data are necessary to successfully back-calculate variables in the presence of significant amounts of noise. In addition, we have studied the sensitivity to sensor characteristics such as detection level and saturation level. It has been applied with Gaussian puff, Gaussian plume, sheared plume and puff models, and SCIPUFF as models for the atmospheric transport and dispersion (ATD). The prediction set denoted "PSU GAUSSIAN" uses Gaussian puff and plume models for ATD, and the prediction set denoted "PSU SCIPUFF" uses SCIPUFF for ATD.

The second method developed at Penn State is the MEFA technique, although the current implementation is for a single entity. It is envisioned as being appropriate for cases where the dispersing eddies are on the scale of the size of the puff or larger, such as in the immediate vicinity of the release. For MEFA, the STE is accomplished by analyzing the evolution of an entity quantity that describes the contaminant distribution, that is, the plume/puff spread. For an instantaneous release, a strictly Lagrangian approach is used with the source information being found by inverting a simple set of equations. In contrast, the formulation for a continuous release cannot adopt this strictly Lagrangian approach because a steady flow of contaminants renders the problem statistically stationary. Therefore, the concentration data are averaged in time, and a hybrid Lagrangian/Eulerian framework is used to analyze the average entity state. It is shown that these entity frameworks are suitable to ascertain source information for a contaminant for dense and sparse sensor grids. An advantage of these algorithms is that no meteorological input is required. Both algorithms were applied to the release and the one with the best prediction used to report the results. The prediction set denoted "PSU MEFA" is based on this method.

References

1. Annunzio, A.J., G.S. Young, and S.E. Haupt, "Combining Methods from Entity and Field Frameworks to Determine the Source Information for a Contaminant," submitted to *Atmospheric Environment*, 2010.

2 Long, K.J., S.E. Haupt, and G.S. Young, "Assessing Sensitivity of Source Term Estimation," *Atmospheric Environment*, 44, 1558-1567, 2010.

3. Haupt, S.E., G.S. Young, and C.T. Allen, "A Genetic Algorithm Method to Assimilate Sensor Data for a Toxic Contaminant Release," *Journal of Computers*, 2, 85-93, 2007.

4. Allen, C.T., G.S. Young, and S.E. Haupt, “Improving Pollutant Source Characterization by Optimizing Meteorological Data with a Genetic Algorithm,” *Atmospheric Environment*, 41, 2283-2289, 2007.
5. Haupt, S.E. G.S. Young, and C.T. Allen, “Validation Of A Receptor/Dispersion Model Coupled With A Genetic Algorithm,” *Journal of Applied Meteorology*, 45, 476–490, 2006.

(This page is intentionally blank.)

Appendix D
Developer Feedback Package Description

The following charts are from the "Directory-Content-and-Keys-to-Charts.ppt" briefing that describes the contents of the developer feedback package distributed to STE developers in September 2009. The developer feedback package contained a root directory and eight main subdirectories corresponding to individual organizations that participated in the evaluation. It contains 1,199 files and 16 folders and occupies 60 MB of disk space.

Summary

- **This appendix describes directory structure and contents of the “Feedback to Developers Package” for Phase I of STE evaluation.**
- **It also describes some keys to provided charts.**

Root Directory Contents

- **At the present, the root directory contains five files and a number of subdirectories for each individual organization that submitted predictions to IDA.**
  - Directory-Contents-and-Keys-to-Charts.ppt is this file.
  - Independent_Variable.xls is an Excel file that contains selected information about actual cases that comprised Phase I. We’re planning to use it for the regression analysis.
    - » Most of the columns are self-explanatory.
    - » Column “# of Puff Realiz > 1” is derived from “# of Realizations” column and is used to distinguish puff cases that have multiple realizations.
      - -1: denotes that the case is based on continuous trial.
      - 0: denotes that a single realization of puff(s) were used in the case.
      - 1: denotes that more than one realizations of puff(s) were used in the case.
      - Sample is shown later.
  - Basic_Intercomparison.{csv, xls, ppt} are files that provide basic model comparisons.
    - » Basic_Itercomparison.csv file is a data file that was imported into Basic_Intercomparison.xls file.
    - » Basic_Intercomparison.xls file contains a number of different charts in separate worksheets that compare model performance.
    - » Basic_Intercomparison.ppt file contains sequence of charts from Basic_Intercomparison.xls file.

Individual Participant Directory

- **Each individual participant in Phase I directory contains a single subdirectory and a number of individual files.**
  - *Individual_Case_csv* subdirectory contains a number of csv files that provide all actual and predicted information for individual cases including location, mass, duration, and start time.
    - » Single file for each predicted case submitted to us.
    - » Sample file shown later.
  - *Location_Plots_{Developer}_{Pred Set}.pdf* is a pdf file containing a series of plots with a single plot for each submitted case showing actual and predicted locations, distance metric, total actual and predicted massed, and maximum concentration color-coded digiPIDs that were used to define each case.
    - » Sample plot is shown later.
  - *Selected_Plots_{Developer}_{Pred Set}.pdf* is a pdf file containing a series of plots that congregate cases according to some selected criteria. It plots all actual and predicted source term locations and provides number of statistics in the legend.
    - » Sample plot is shown later.

Individual Participant Directory (Cont'd)

- *Predicted_Locations_Mass_Stat_{Developer}_{Pred Set}.xls* is an Excel file that provides a number of charts comparing a particular set of predictions in terms of a select subset of conditions.
  - » Primary conditions include "4 vs .16," "Operational vs. Close-In Met," "Number of Sources," "Day vs. Night."
    - Secondary conditions are varied.
  - » Both "total" mass and "Average Distance" metrics are used.
  - » Sample chart is shown later.
- *Predicted_Locations_Stat_dump_{Developer}_{Pred Set}.csv* is an ASCII file that provides statistics for individual algorithm performance based on a large set of conditions.
  - » Both total mass and "averaged distance" metrics are provided.
    - For "total mass" statistic "Mean A_Mass" and "Median A_Mass" denotes actual release values and "Mean P_Mass" and "Median P_Mass" denotes predicted values.
  - » Both "mean" and "median" are provided.
  - » Small subset of this file is used to create charts in the Excel file described in the previous bullet.
  - » Sample file is shown later.
- *Actual_vs_Observed_Release_Type_Comparison_{Developer}_{Pred Set}.csv* is an ASCII file used for debugging purposes. We decided to include it here since we expect that some developers might find it useful. It provides a single Worksheet with limited source term information for each individual case (both actual and predicted).
  - » Individual csv files inside *Individual_Cases_csv* subdirectory contain more information.
  - » Provided information includes release duration, release type, number of locations, and number of realizations at each location.
  - » Sample file is shown later.

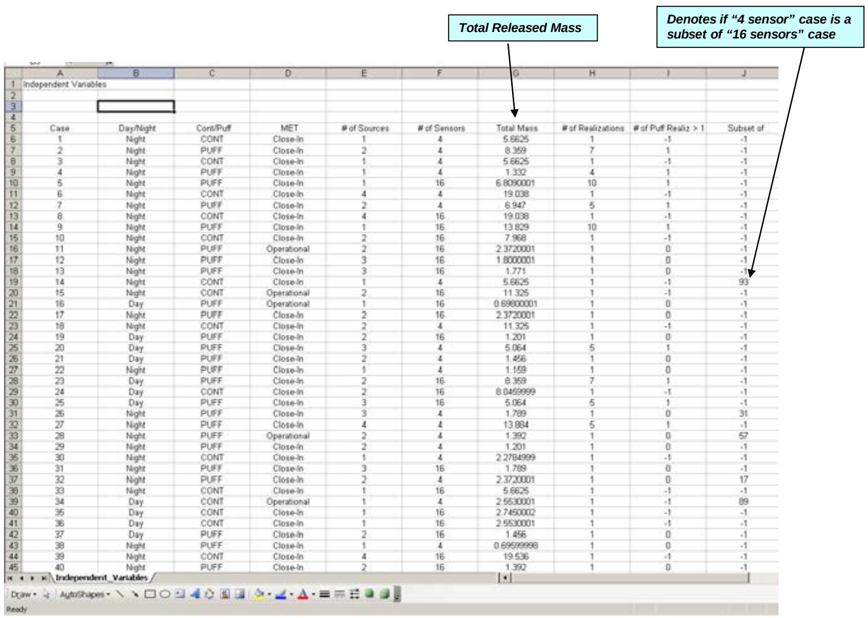
Individual Participant Directory (Cont’d)

– *Dependent_Variables_{Developer}_{Pred Set}.csv* is an ASCII file that contains “Distance” metrics and total predicted mass for each set of individual predictions. We’re planning to use it for the regression analysis.
  » Sample file is shown later.

– *Triple_Bar_Chart_*_{Developer}_{Pred Set}_Comparable.png* are four bar-charts that were distributed earlier. The description and samples of these bar charts are shown later in this appendix.

Sample Files and Charts

Independent_Variable.xls

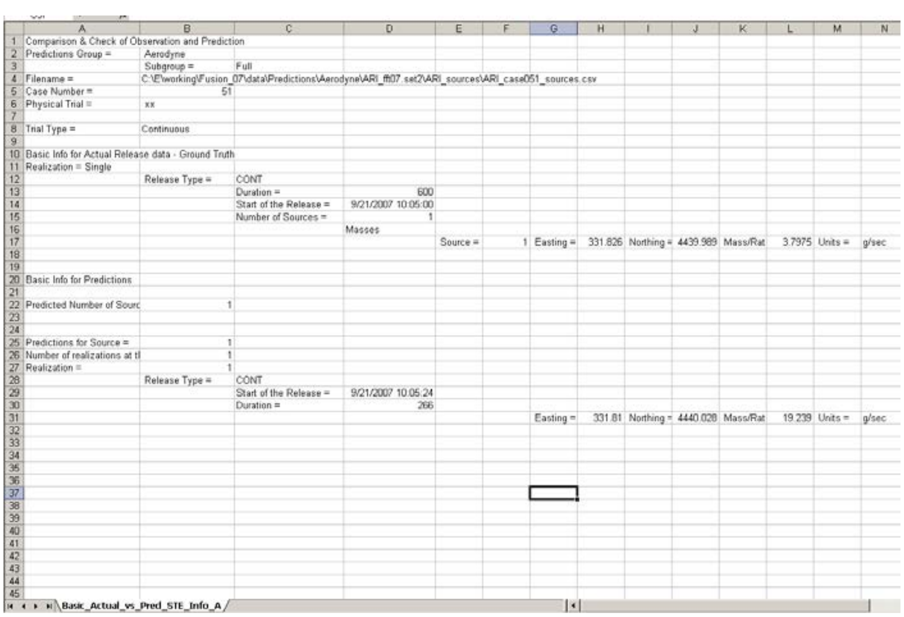


Independent Variables

Case	Day/Night	Cont/Puff	MET	# of Sources	# of Sensors	Total Mass	# of Realizations	# of Puff Realiz > 1	Subset of
1	Night	CONT	Close-In	1	4	5.6625	1	-1	-1
2	Night	PUFF	Close-In	2	4	8.359	7	1	-1
3	Night	CONT	Close-In	1	4	5.6625	1	-1	-1
4	Night	PUFF	Close-In	1	4	1.332	4	1	-1
5	Night	PUFF	Close-In	1	16	6.8090001	10	1	-1
6	Night	CONT	Close-In	4	4	19.038	1	-1	-1
7	Night	PUFF	Close-In	2	4	6.947	5	1	-1
8	Night	CONT	Close-In	4	16	19.038	1	-1	-1
9	Night	PUFF	Close-In	1	16	13.829	10	1	-1
10	Night	CONT	Close-In	2	16	7.968	1	-1	-1
11	Night	PUFF	Operational	2	16	2.3720001	1	0	-1
12	Night	PUFF	Close-In	3	16	1.8000001	1	0	-1
13	Night	PUFF	Close-In	3	16	1.771	1	0	-1
14	Night	CONT	Close-In	1	4	5.6625	1	-1	93
15	Night	CONT	Operational	2	16	11.325	1	-1	-1
16	Day	PUFF	Operational	1	16	0.69800001	1	0	-1
17	Night	PUFF	Close-In	2	16	2.3720001	1	0	-1
18	Night	CONT	Close-In	2	4	11.325	1	-1	-1
19	Day	PUFF	Close-In	2	16	1.201	1	0	-1
20	Day	PUFF	Close-In	3	4	5.064	5	1	-1
21	Day	PUFF	Close-In	2	4	1.456	1	0	-1
22	Night	PUFF	Close-In	1	4	1.159	1	0	-1
23	Day	PUFF	Close-In	2	16	8.359	7	1	-1
24	Day	CONT	Close-In	2	16	8.0469999	1	-1	-1
25	Day	PUFF	Close-In	3	16	5.064	5	1	-1
26	Night	PUFF	Close-In	3	4	1.789	1	0	31
27	Night	PUFF	Close-In	4	4	13.884	5	1	-1
28	Night	PUFF	Operational	2	4	1.392	1	0	57
29	Night	PUFF	Close-In	2	4	1.201	1	0	-1
30	Night	CONT	Close-In	1	4	2.2784999	1	-1	-1
31	Night	PUFF	Close-In	3	16	1.789	1	0	-1
32	Night	PUFF	Close-In	2	4	2.3720001	1	0	17
33	Night	CONT	Close-In	1	16	5.6625	1	-1	-1
34	Day	CONT	Operational	1	4	2.5530001	1	-1	89
35	Day	CONT	Close-In	1	16	2.7450002	1	-1	-1
36	Day	CONT	Close-In	1	16	2.5530001	1	-1	-1
37	Day	PUFF	Close-In	2	16	1.456	1	0	-1
38	Night	PUFF	Close-In	1	4	0.69599998	1	0	-1
39	Night	CONT	Close-In	4	16	19.536	1	-1	-1
40	Night	PUFF	Close-In	2	16	1.392	1	0	-1

Individual_Case_csv Subdirectory Sample

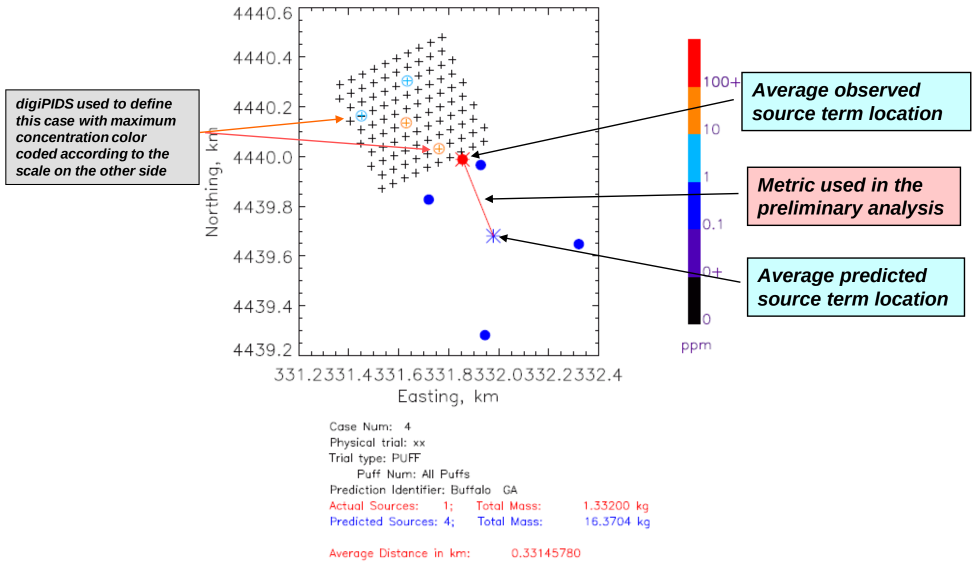
Basic_Actual_vs_Pred_STE_Info_Aerodyne_Full_Case_051.csv



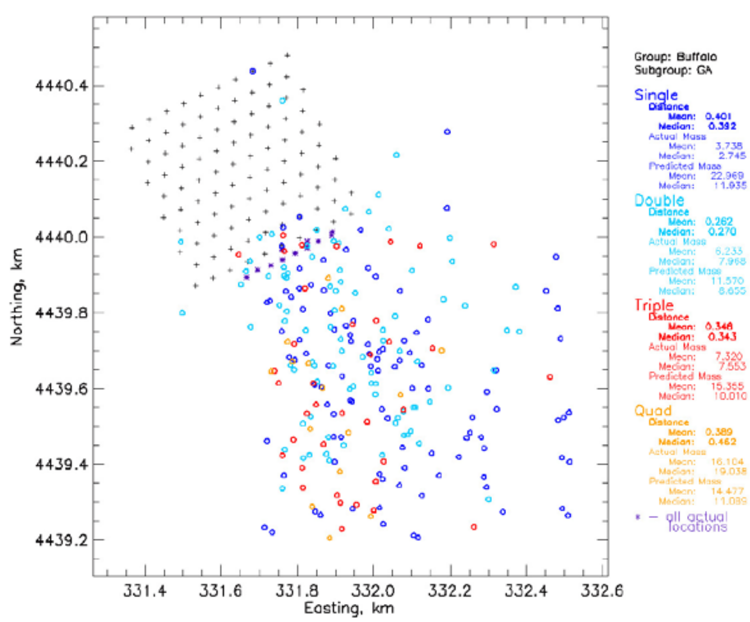
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Comparison & Check of Observation and Prediction													
2	Predictions Group =	Aerodyne												
3		Subgroup =	Full											
4	Filename =	C:\E\working\Fusion_07\data\Predictions\Aerodyne\ARI_ff07.set2\ARI_sources\ARI_case051_sources.csv												
5	Case Number =	51												
6	Physical Trial =	xx												
8	Trial Type =	Continuous												
10	Basic Info for Actual Release data - Ground Truth													
11	Realization = Single													
12		Release Type =	CONT											
13			Duration =	600										
14			Start of the Release =	9/21/2007 10:05:00										
15			Number of Sources =	1										
16				Masses										
17					Source =	1	Easting =	331.826	Northing =	4439.989	Mass/Rat	3.7975	Units =	g/sec
20	Basic Info for Predictions													
22	Predicted Number of Sourc	1												
25	Predictions for Source =	1												
26	Number of realizations at tl	1												
27	Realization =	1												
28		Release Type =	CONT											
29			Start of the Release =	9/21/2007 10:05:24										
30			Duration =	266										
31							Easting =	331.81	Northing =	4440.028	Mass/Rat	19.239	Units =	g/sec

Metric Used in the Preliminary Analysis

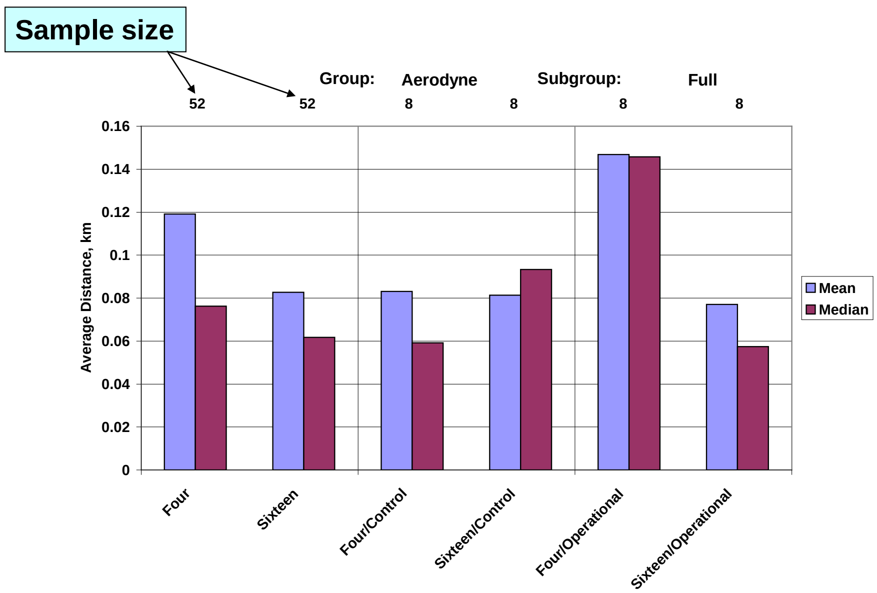
Sample Plot in *Location_Plots_Buffalo_GA.pdf*



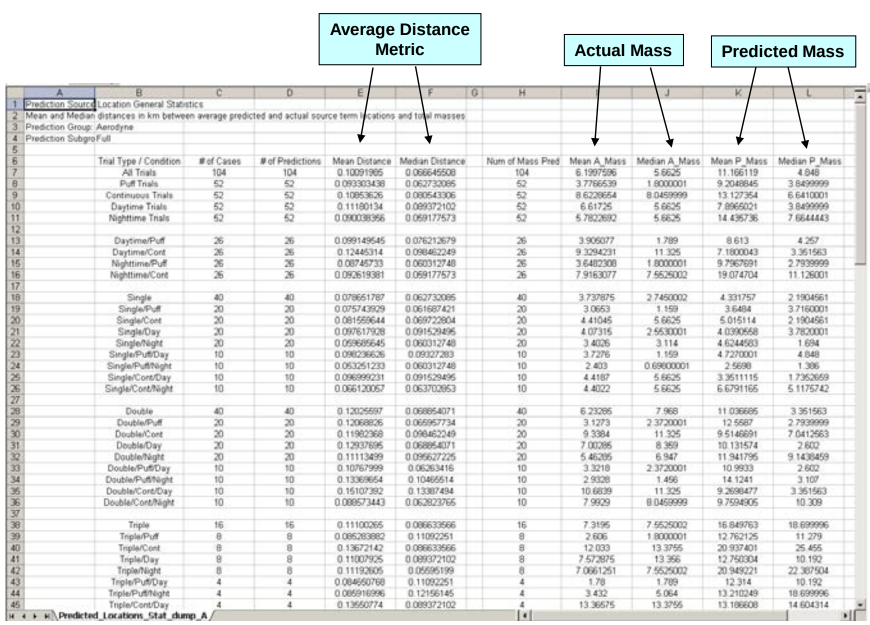
Sample Plot in *Selected_Plots_Buffalo_GA.pdf*



Sample Chart in *Predicted_Locations_Mass_Stat_Aerodyne_Full.xls*



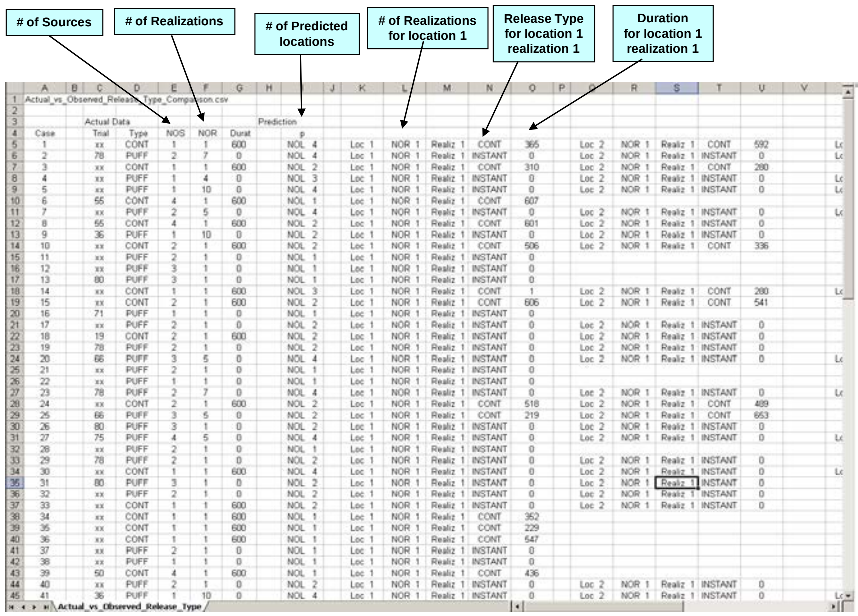
Sample *Predicted_Locations_Stat_dump_Aerodyne_Full.csv*



Prediction Source Location General Statistics
Mean and Median distances in km between average predicted and actual source term locations and total masses
Prediction Group: Aerodyne
Prediction Subgro Full

Trial Type / Condition	# of Cases	# of Predictions	Mean Distance	Median Distance	Num of Mass Pred	Mean A_Mass	Median A_Mass	Mean P_Mass	Median P_Mass
All Trials	104	104	0.10091985	0.066645508	104	6.1997596	5.6625	11.166119	4.848
Puff Trials	52	52	0.093303438	0.062732085	52	3.7766539	1.8000001	9.2048845	3.8499999
Continuous Trials	52	52	0.10853626	0.080543306	52	8.6228654	8.0459999	13.127354	6.6410001
Daytime Trials	52	52	0.11180134	0.089372102	52	6.61725	5.6625	7.8965021	3.8499999
Nighttime Trials	52	52	0.090038356	0.059177573	52	5.7822692	5.6625	14.435736	7.6644443
Daytime/Puff	26	26	0.099149545	0.076212679	26	3.905077	1.789	8.613	4.257
Daytime/Cont	26	26	0.12445314	0.098462249	26	9.3294231	11.325	7.1800043	3.351563
Nighttime/Puff	26	26	0.08745733	0.060312748	26	3.6482308	1.8000001	9.7967691	2.7939999
Nighttime/Cont	26	26	0.092619381	0.059177573	26	7.9163077	7.5525002	19.074704	11.126001
Single	40	40	0.078651787	0.062732085	40	3.737875	2.7450002	4.331757	2.1904561
Single/Puff	20	20	0.075743929	0.061687421	20	3.0653	1.159	3.6484	3.7160001
Single/Cont	20	20	0.081559644	0.069722804	20	4.41045	5.6625	5.015114	2.1904561
Single/Day	20	20	0.097617928	0.091529496	20	4.07315	2.5530001	4.0390558	3.7820001
Single/Night	20	20	0.059685645	0.060312748	20	3.4026	3.114	4.6244583	1.694
Single/Puff/Day	10	10	0.098236626	0.09327283	10	3.7276	1.159	4.7270001	4.848
Single/Puff/Night	10	10	0.053251233	0.060312748	10	2.403	0.69800001	2.5698	1.386
Single/Cont/Day	10	10	0.096999231	0.091529496	10	4.4187	5.6625	3.3511115	1.7352659
Single/Cont/Night	10	10	0.066120057	0.063702853	10	4.4022	5.6625	6.6791165	5.1175742
Double	40	40	0.12025597	0.068854071	40	6.23285	7.968	11.036685	3.351563
Double/Puff	20	20	0.12068826	0.065957734	20	3.1273	2.3720001	12.5587	2.7939999
Double/Cont	20	20	0.11982368	0.098462249	20	9.3384	11.325	9.5146691	7.0412563
Double/Day	20	20	0.12937695	0.068854071	20	7.00285	8.369	10.131574	2.602
Double/Night	20	20	0.11113499	0.095627225	20	5.46285	6.947	11.941795	9.1438459
Double/Puff/Day	10	10	0.10767999	0.06263416	10	3.3218	2.3720001	10.9933	2.602
Double/Puff/Night	10	10	0.13369654	0.10465514	10	2.9328	1.456	14.1241	3.107
Double/Cont/Day	10	10	0.15107392	0.13387494	10	10.6839	11.325	9.2698477	3.351563
Double/Cont/Night	10	10	0.088573443	0.062823765	10	7.9929	8.0459999	9.7594905	10.309
Triple	16	16	0.11100265	0.086633566	16	7.3195	7.5525002	16.849763	18.699996
Triple/Puff	8	8	0.085283882	0.11092251	8	2.606	1.8000001	12.762125	11.279
Triple/Cont	8	8	0.13672142	0.086633566	8	12.033	13.3755	20.937401	25.455
Triple/Day	8	8	0.11007925	0.089372102	8	7.572875	13.356	12.750304	10.192
Triple/Night	8	8	0.11192605	0.05595199	8	7.0661251	7.5525002	20.949221	22.387504
Triple/Puff/Day	4	4	0.084650768	0.11092251	4	1.78	1.789	12.314	10.192
Triple/Puff/Night	4	4	0.085916996	0.12156145	4	3.432	5.064	13.210249	18.699996
Triple/Cont/Day	4	4	0.13550774	0.089372102	4	13.36575	13.3755	13.186608	14.604314

Sample *Actual_vs_Observed_Release_Type_Comparison_Aerodyne_Full.csv*

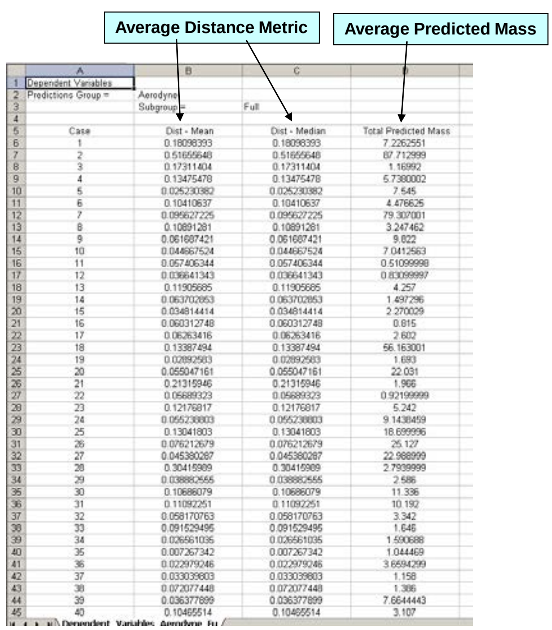


Actual_vs_Observed_Release_Type_Comparison.csv

	Actual Data					Prediction											
Case	Trial	Type	NOS	NOR	Durat	p											
1	xx	CONT	1	1	600	NOL 4	Loc 1	NOR 1	Realiz 1	CONT	365	Loc 2	NOR 1	Realiz 1	CONT	592	Lc
2	78	PUFF	2	7	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
3	xx	CONT	1	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	310	Loc 2	NOR 1	Realiz 1	CONT	280	
4	xx	PUFF	1	4	0	NOL 3	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
5	xx	PUFF	1	10	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
6	55	CONT	4	1	600	NOL 1	Loc 1	NOR 1	Realiz 1	CONT	607						
7	xx	PUFF	2	5	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
8	55	CONT	4	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	601	Loc 2	NOR 1	Realiz 1	INSTANT	0	
9	36	PUFF	1	10	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
10	xx	CONT	2	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	506	Loc 2	NOR 1	Realiz 1	CONT	336	
11	xx	PUFF	2	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
12	xx	PUFF	3	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
13	80	PUFF	3	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
14	xx	CONT	1	1	600	NOL 3	Loc 1	NOR 1	Realiz 1	CONT	1	Loc 2	NOR 1	Realiz 1	CONT	280	Lc
15	xx	CONT	2	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	606	Loc 2	NOR 1	Realiz 1	CONT	541	
16	71	PUFF	1	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
17	xx	PUFF	2	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
18	19	CONT	2	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
19	78	PUFF	2	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
20	66	PUFF	3	5	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
21	xx	PUFF	2	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
22	xx	PUFF	1	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
23	78	PUFF	2	7	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
24	xx	CONT	2	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	518	Loc 2	NOR 1	Realiz 1	CONT	489	
25	66	PUFF	3	5	0	NOL 2	Loc 1	NOR 1	Realiz 1	CONT	219	Loc 2	NOR 1	Realiz 1	CONT	653	
26	80	PUFF	3	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
27	75	PUFF	4	5	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
28	xx	PUFF	2	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
29	78	PUFF	2	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
30	xx	CONT	1	1	600	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc
31	80	PUFF	3	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
32	xx	PUFF	2	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
33	xx	CONT	1	1	600	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
34	xx	CONT	1	1	600	NOL 1	Loc 1	NOR 1	Realiz 1	CONT	352						
35	xx	CONT	1	1	600	NOL 1	Loc 1	NOR 1	Realiz 1	CONT	229						
36	xx	CONT	1	1	600	NOL 1	Loc 1	NOR 1	Realiz 1	CONT	547						
37	xx	PUFF	2	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
38	xx	PUFF	1	1	0	NOL 1	Loc 1	NOR 1	Realiz 1	INSTANT	0						
39	50	CONT	4	1	600	NOL 1	Loc 1	NOR 1	Realiz 1	CONT	436						
40	xx	PUFF	2	1	0	NOL 2	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	
41	36	PUFF	1	10	0	NOL 4	Loc 1	NOR 1	Realiz 1	INSTANT	0	Loc 2	NOR 1	Realiz 1	INSTANT	0	Lc

Actual_vs_Observed_Release_Type

Sample *Dependent_Variables_Aerodyne_Full.csv*



Dependent Variables

Predictions Group = Aerodyne

Subgroup = Full

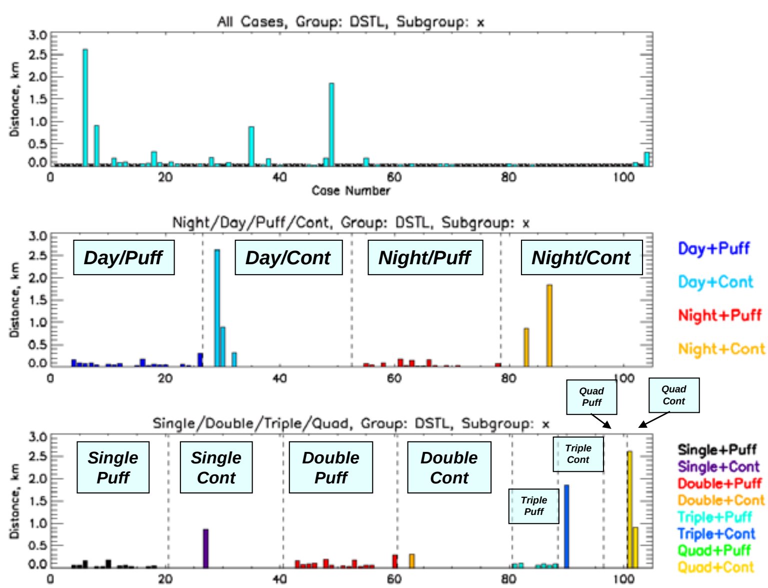
Case	Dist - Mean	Dist - Median	Total Predicted Mass
1	0.18098393	0.18098393	7.2262551
2	0.51655648	0.51655648	87.712999
3	0.17311404	0.17311404	1.16992
4	0.13475478	0.13475478	5.7380002
5	0.025230382	0.025230382	7.545
6	0.10410637	0.10410637	4.476625
7	0.095627225	0.095627225	79.307001
8	0.10891281	0.10891281	3.247462
9	0.061687421	0.061687421	9.822
10	0.044667524	0.044667524	7.0412563
11	0.057406344	0.057406344	0.51099998
12	0.036641343	0.036641343	0.83099997
13	0.11905685	0.11905685	4.257
14	0.063702853	0.063702853	1.497296
15	0.034814414	0.034814414	2.270029
16	0.060312748	0.060312748	0.815
17	0.06263416	0.06263416	2.602
18	0.13387494	0.13387494	56.163001
19	0.02892583	0.02892583	1.693
20	0.055047161	0.055047161	22.031
21	0.21315946	0.21315946	1.966
22	0.05689323	0.05689323	0.92199999
23	0.12176817	0.12176817	5.242
24	0.055238803	0.055238803	9.1438459
25	0.13041803	0.13041803	18.699996
26	0.076212679	0.076212679	25.127
27	0.045380287	0.045380287	22.988999
28	0.30415989	0.30415989	2.7939999
29	0.038882555	0.038882555	2.586
30	0.10686079	0.10686079	11.336
31	0.11092251	0.11092251	10.192
32	0.058170763	0.058170763	3.342
33	0.091529495	0.091529495	1.646
34	0.026561035	0.026561035	1.590688
35	0.007267342	0.007267342	1.044469
36	0.022979246	0.022979246	3.6594299
37	0.033039803	0.033039803	1.158
38	0.072077448	0.072077448	1.386
39	0.036377899	0.036377899	7.6644443
40	0.10465514	0.10465514	3.107

Dependent_Variables_Aerodyne_Fu

Sample Triple-Bar Charts
Note to Triple-Bar Charts

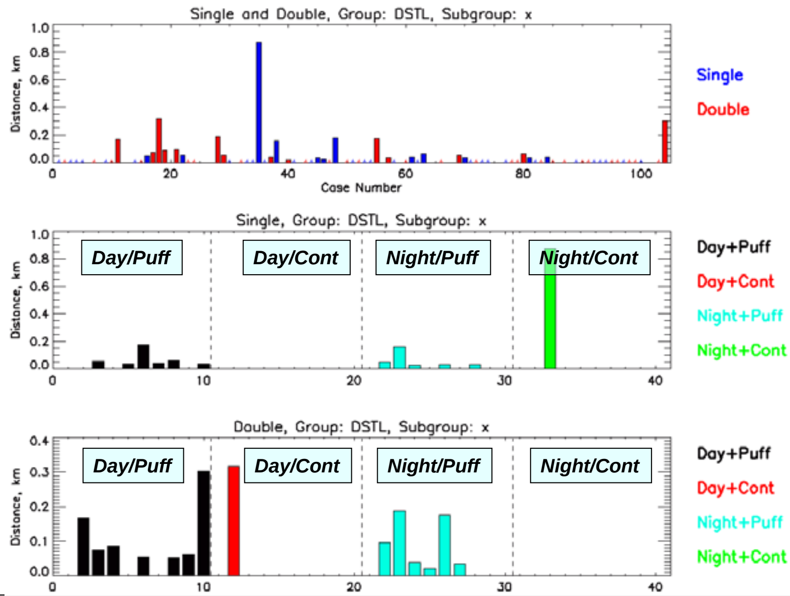
- **Bottom two panels in bar charts in the next two charts are slightly modified from the version of the bar charts presented at TP9 meeting.**
- **Horizontal axes are divided into "blocks" corresponding to criteria of interest, and each case that was distributed has a "fixed" position within the block.**
- **Modified bar charts could be used for individual cases inter-comparison between different model predictions.**
  - Unlike bar charts that were presented at TP9 meeting.

Typical "Distance Charts"
DSTL Predictions (Linear), All Cases



Typical "Distance Charts"

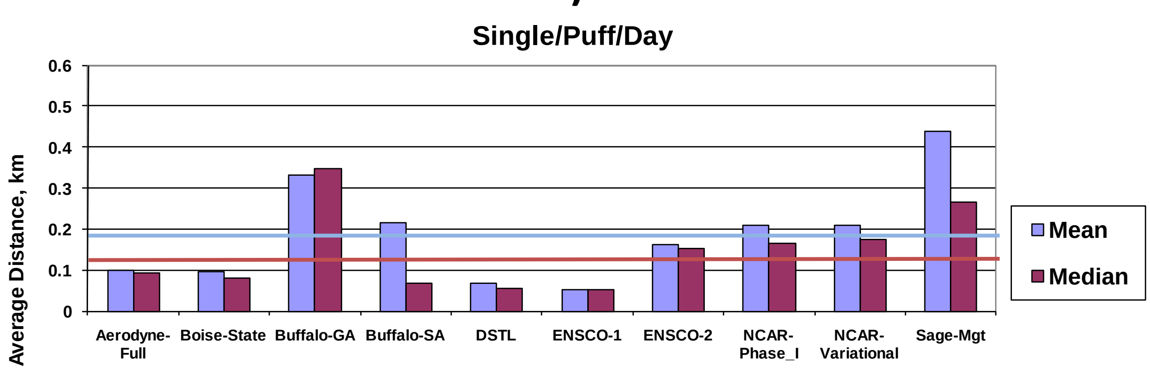
DSTL Predictions (Linear), Single and Double



Appendix E
Additional Plots for Miss Distance Intercomparison

Figures E-1 through E-4 compare the performance of individual algorithms using the averaged and median miss distance metric, where the average or median is taken over all predicted cases in the subgroup. Each figure consists of two parts: a) depicts daytime and b) depicts nighttime algorithm performance. The light blue line shows the median of the “mean” distance; the purple line shows the median of the “median” distance. Figures E-5 through E-8 depict the median miss distance metric, where the median is taken over all predicted cases in the subgroup with different breakdowns of individual subgroups for easier comparisons of algorithm performance.

a)



b)

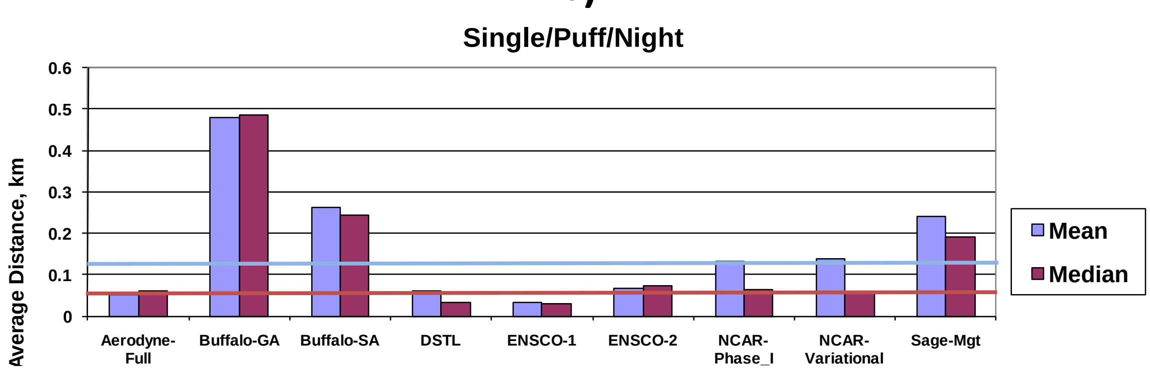


Figure E-1. STE Algorithm Comparison Based On Single-Source/Instantaneous Releases

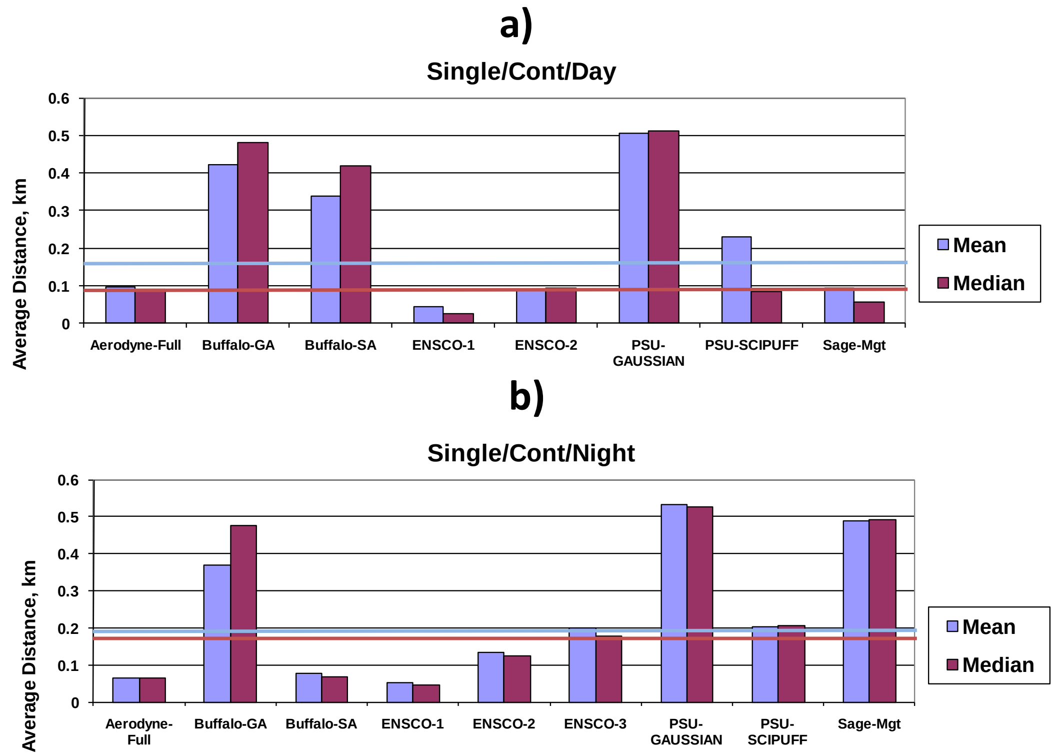


Figure E-2. STE Algorithm Comparison Based On Single-Source/Continuous Releases

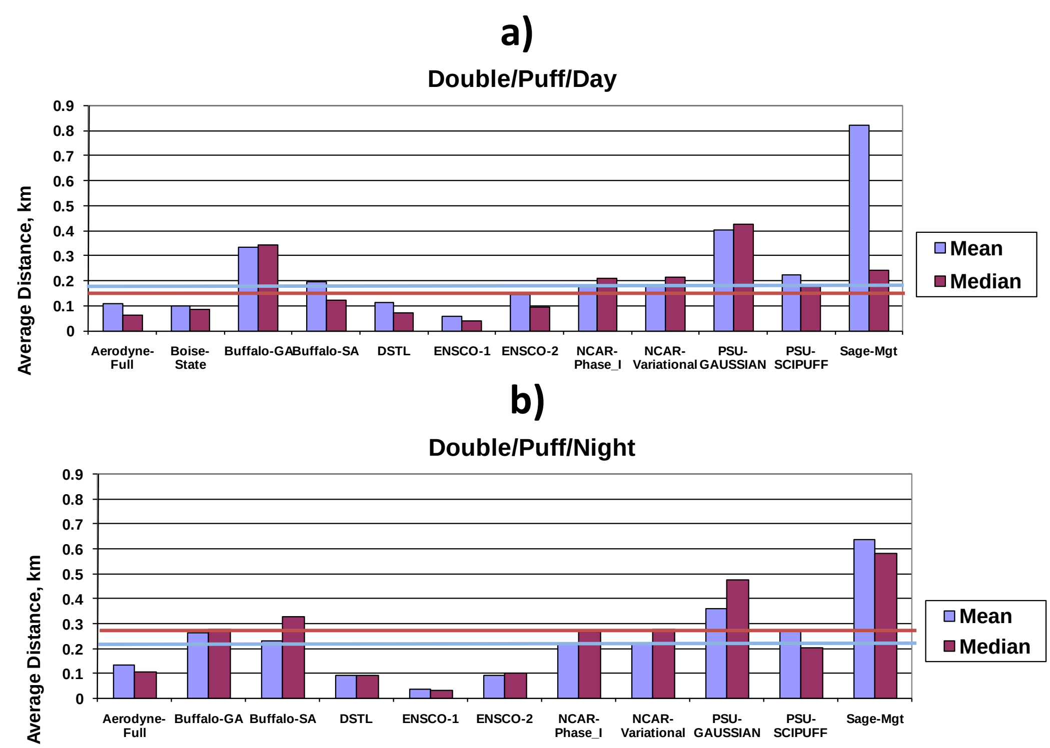
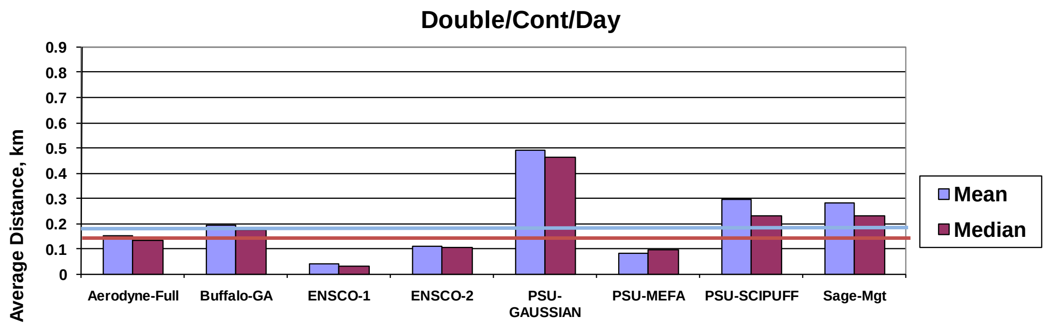


Figure E-3. STE algorithm Comparison Based on Double-Source/Instantaneous Releases

a)



b)

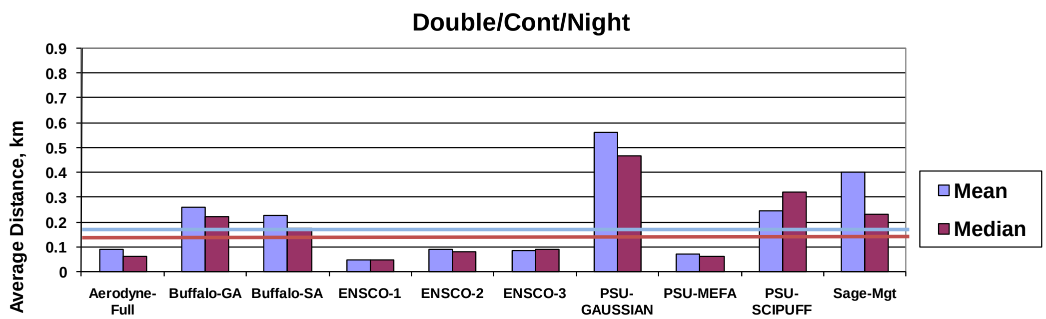


Figure E-4. STE Algorithm Comparison Based on Double-Source/Continuous Releases

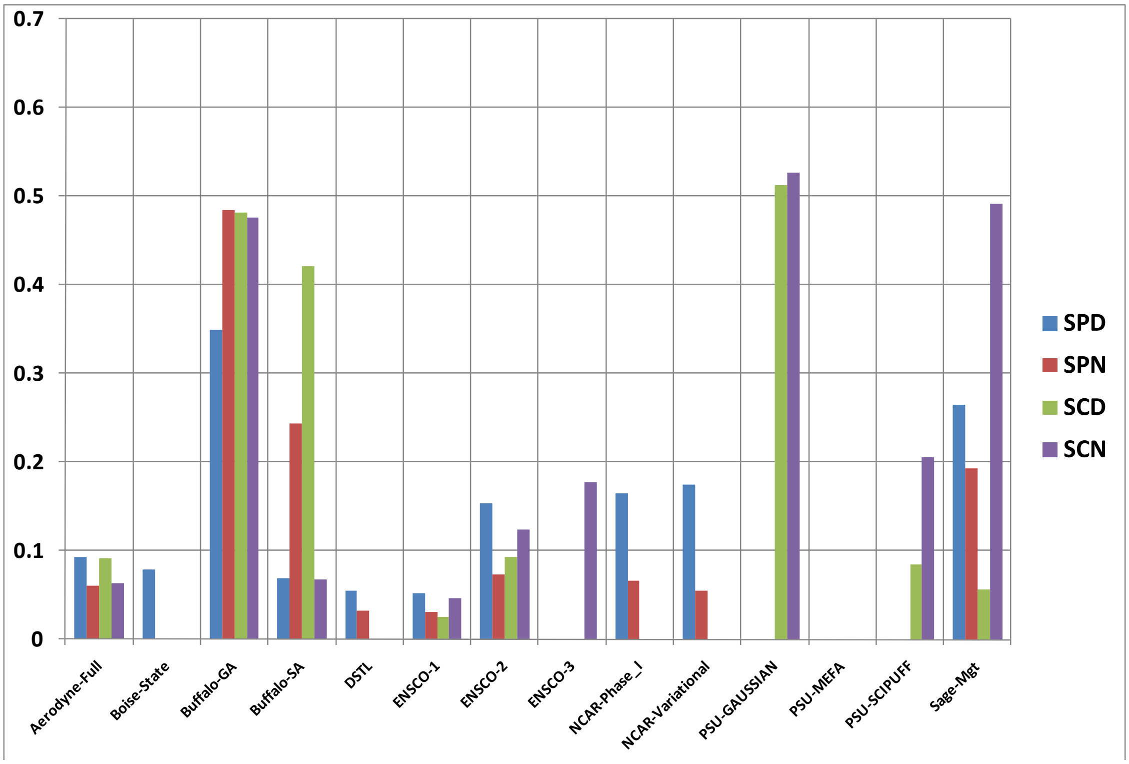


Figure E-5. STE Algorithm Comparison Based on Single-Source Releases

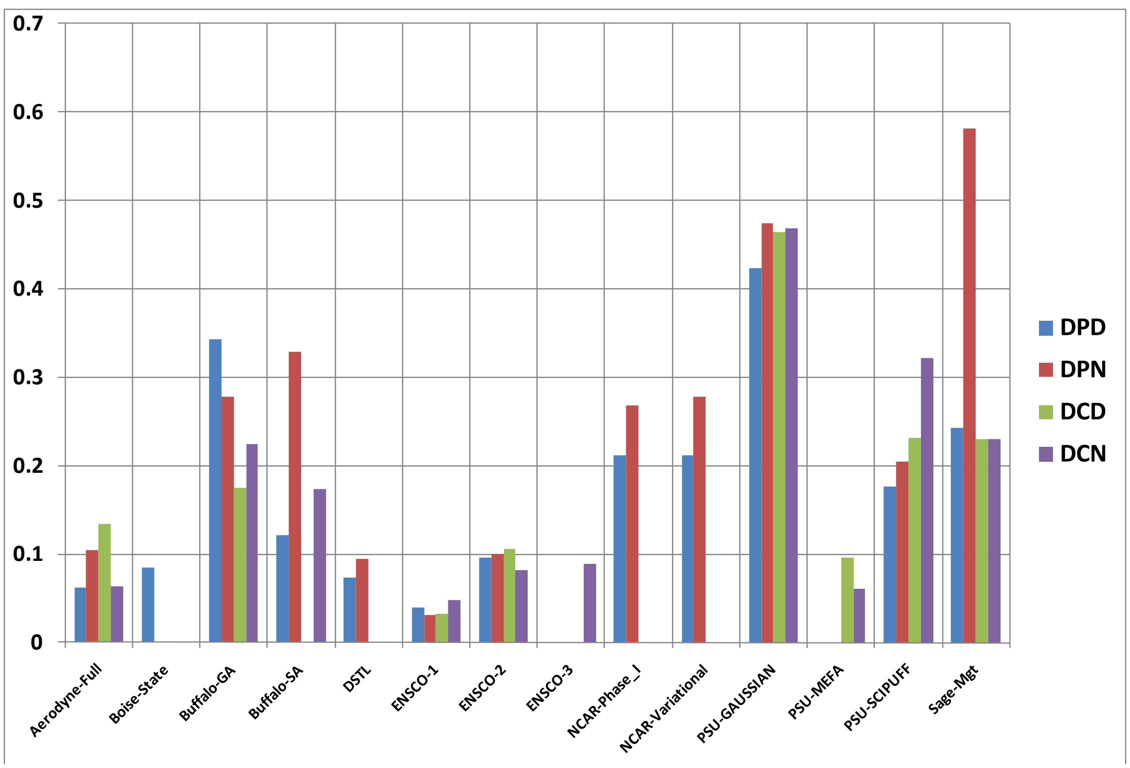


Figure E-6. STE Algorithm Comparison Based on Double-Source Releases

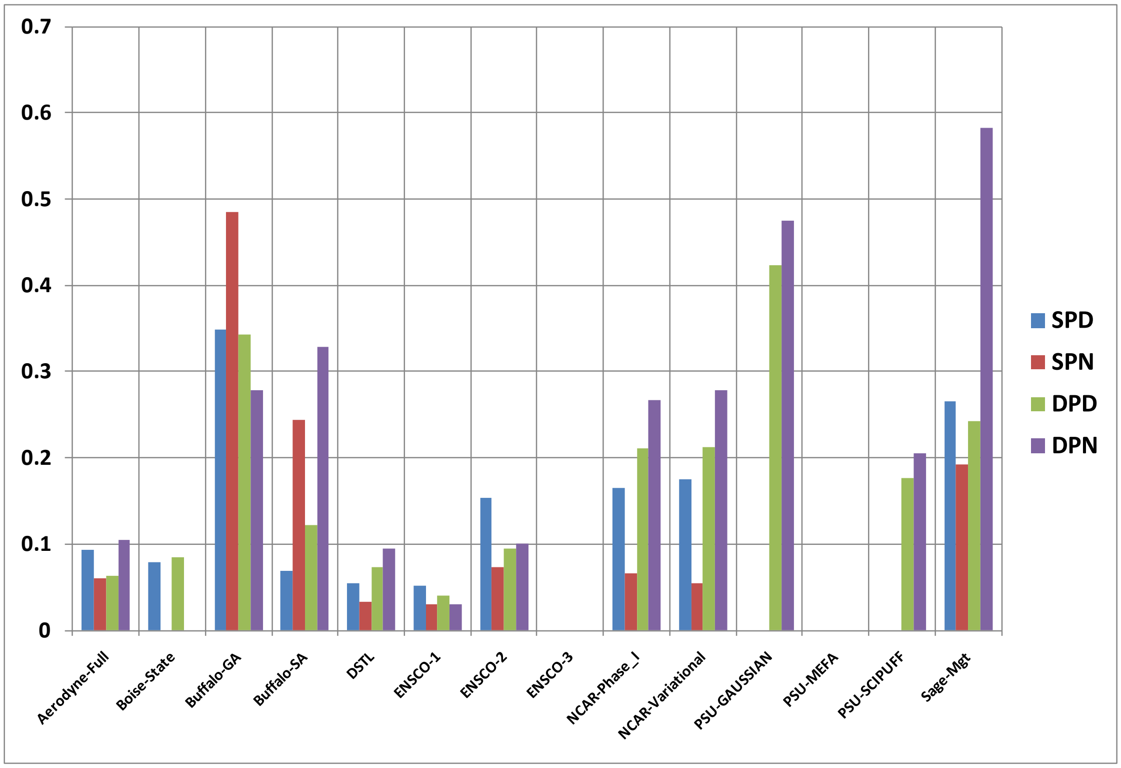


Figure E-7. STE Algorithm Comparison Based on Instantaneous Releases

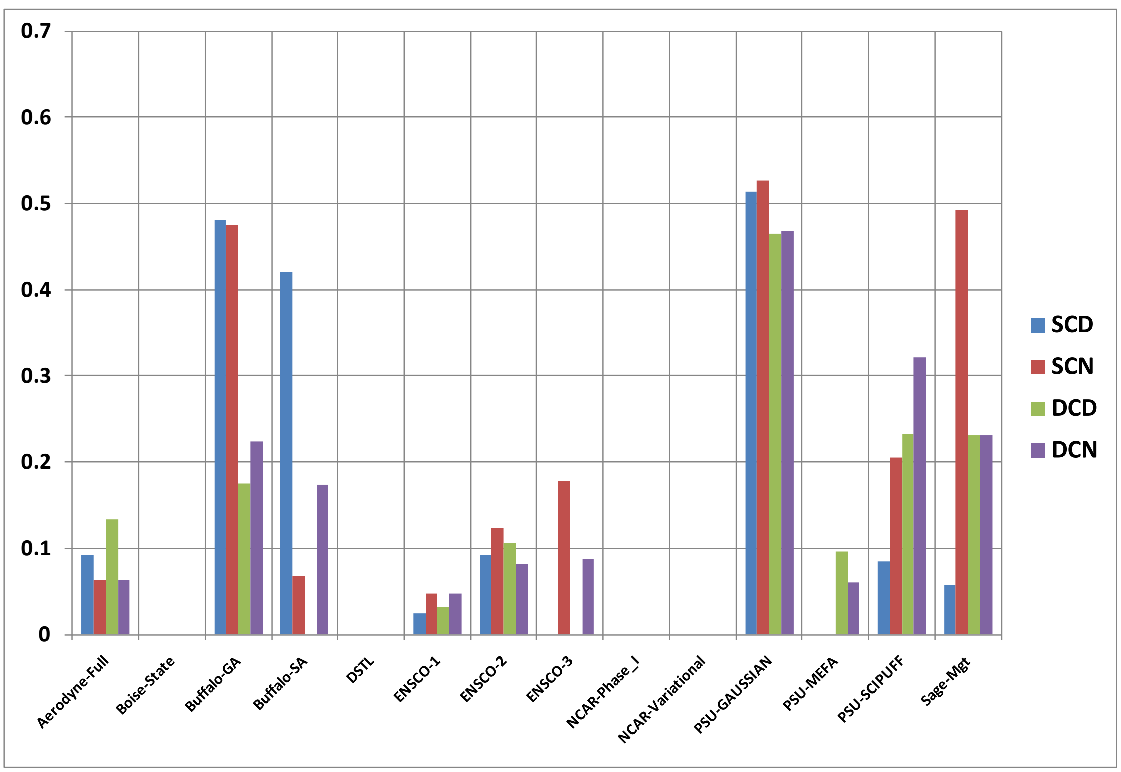


Figure E-8. STE Algorithm Comparison Based on Continuous Releases

(This page is intentionally blank.)

Appendix F
Linear Regression

A. Description

This section describes the use of stepwise and backward linear regression for the examination of source term estimation algorithms. In general terms, this effort was an attempt to determine which of the underlying factors, such as diurnal condition, number of release sources, type of release, and several other independent variables, had the greatest effect on the estimation of the mass ratio (the ratio of reported to actual mass) or distance. Standard linear regression determines a set of coefficients for independent variables that yield the smallest sum of squares of residuals (differences between observed data and their linear approximation). Stepwise and backward regression each perform this "least squares fit" but additionally attempt to include in the regression equation only those independent variables that substantially reduce this sum of squares. In this sense, they are more parsimonious than standard regression.

Stepwise regression begins by selecting the independent variable that is most highly correlated with the dependent variable. It performs a regression (i.e., selects a constant term and a coefficient that yield a "least squares fit" to the data) of this variable against the dependent variable. It then selects from the remaining independent variables the one whose partial correlation with the dependent variable (that is, whose correlation after controlling for the effect of the first independent variable) is the highest. The sum of squares associated with this variable is tested with a "partial F-test." If significant, this variable enters the regression equation.

Next, after selecting this second variable, it reexamines the effect of the first independent variable. That is, the first variable is treated as though it were the last variable to enter the regression equation. In this role reversal, the reduction in the sum of squares of the residuals due to the first variable is computed. If this reduction in the sum of squares is not significant (as determined by the appropriate "partial F-test"), the first variable is removed.

The entire process is continued by selecting independent variables with high partial correlations, then treating the previously selected variables as though they were the last to enter the regression equation and eliminating those that do not significantly reduce the sum of squares of the residuals.

Backward regression, like stepwise, is selective in its choice of independent variables. However, it differs substantially from its sister technique by treating every independent variable as though it were the last to enter the regression equation (in other words, there is no “entrance” qualification). The contribution of each in reducing the sum of squares is tested sequentially (with the “partial F-test” mentioned above). Those variables that fall below a prescribed standard are eliminated.

Thus, roughly speaking, stepwise regression tests each independent variable to determine whether it should enter the regression equation, and again, if it should remain in the equation after others are admitted. Backward regression initially treats all variables as belonging to the equation, then eliminates those whose contribution is substandard. For reference please see References. F-1 and F-2.

B. Summary of the Results

The results for stepwise and backward regressions are summarized in Tables F-1 and F-2. Each table is divided into two sections, one for each dependent variable. Each section contains the proportion of variance explained by regression (adjusted R^2), independent variables selected by backward regression, standard coefficient for that variable, unstandardized coefficient, and significance level. To simplify viewing these tables, the colored background in the table entries is coded according to which independent variable is called by the particular significant factor. All computations were performed using SPSS 15.0 [F-3] with a removal criterion of 10-percent significance as determined by the appropriate partial F-test.

The regression outcomes were ranked in decreasing order of their respective adjusted R^2. It is equal to the proportion of the variance in the observed data that can be “explained” by regression, modified by the number of independent variables [F-2]. The adjusted R^2, which determined the ordering, takes into account the number of variables in the model and is equal to $1 - (1 - R^2)\,(n - 1)/(n - p - 1)$, where p is the number of independent variables in regression equation and n is the number of observations. The point of using the adjusted R^2 is to force models to be economical by penalizing excessive numbers of independent variables. This is in contrast to the (unadjusted) R^2, which increases with the number of independent variables.

Table F-1. Table of Significant Factors for Backward Regression

model	dependent	R2	significant factor	significant factor	significant factor
ENSCO 3	Mass Ratio	0.379	Puff Real (0.51, 2.49. 0)	Sources (-0.447, -1.9, 0.001)	
Buffalo SA	Mass Ratio	0.273	Sources (-0.348, -0.723, 0.002)	Met Num (0.235, 0.632, 0.031)	Diurnal (0.231, 0.508, 0.029)
DSTL	Mass Ratio	0.254	Puff Real (-0.567, -287.1, 0.001)	Sources (-0.376, -75.9, 0.026)	
ENSCO 2	Mass Ratio	0.221	Puff Real (0.37, 1.3, 0.0)	Sources (-0.32, -0.93,0)	Sensors (0.17, 0.074, 0.06)
PSA Gaussian	Mass Ratio	0.209	Puff Real (0.46, 0.059, 0.01)	SourceS (-0.407, -0.037, 0.02)	
PSU SCIPUFF	Mass Ratio	0.203	Sources (-0.5, -0.011, 0.035)		
Buffalo GA	Mass Ratio	0.172	Sources (-0.365, -2.376, 0)	Puff Real (0.183, 1.417, 0.044)	Diurnal (0.177, 1.224, 0.051)
ENSCO 1	Mass Ratio	0.15	Puff Real (0.398, 14.64, 0)		
Aerodyne	Mass Ratio	0.096	Puff Real (0.262, 0.852, 0.006)	Sensors (-0.212, -0.089, 0.026)	
NCAR Phase I	Mass Ratio	0	constant		
NCAR Variation	Mass Ratio	0			
SAGE Mgt August	Mass Ratio	0			
Boise State	Mass Ratio		NO DATA		
PSU MEFA	Mass Ratio		NO DATA		
model	dependent	R2	significant factor	significant factor	significant factor
DSTL	Mean	0.67	Puff Real (-0.725, -1.105, 0)	Sources (0.212,0.129, 0.056)	
NCAR Phase I	Mean	0.266	Sources (0.534, 0.09, 0.001)		
NCAR Variation	Mean	0.204	Sources (0.475, 0.09, 0.003)		
ENSCO 3	Mean	0.148	Sources (-0.366, -0.031, 0.015)	Sensors (0.258, 0.003, 0.08)	
PSA Gaussian	Mean	0.102	Sources(0.306, 0.055, 0.029)	Puff Real (-0.254, -0.057, 0.069)	
SAGE Mgt August	Mean	0.083	Sources (0.303, 0.204, 0.002)		
ENSCO 1	Mean	0.043	Met Num (0.228, 0.009, 0.021)		
ENSCO 2	Mean	0.04	Sensors (-0.173, -0.002, 0.076)	Met Num (0.169, 0.017, 0.083)	
Aerodyne	Mean	0.033	Sensors (-0.206, -0.003, 0.036)		
Boise State	Mean	0	constant		
Buffalo GA	Mean	0	constant		
Buffalo SA	Mean	0			
PSU MEFA	Mean	0	constant		
PSU SCIPUFF	Mean	0	constant		

Table F-2. Table of Significant Factors for Stepwise Regression

model	dependent	R2	significant factor	significant factor	significant factor
ENSCO 3	Mass Ratio	0.379	Puff Real (0.51, 2.49. 0)	Sources (-0.447, -1.9, 0.001)	
Buffalo SA	Mass Ratio	0.273	Sources (-0.348, -0.723, 0.002)	Met Num (0.235, 0.632, 0.031)	Diurnal (0.231, 0.508, 0.029)
DYSTL	Mass Ratio	0.254	Puff Real (-0.567, -287.1, 0.001)	Sources (-0.376, -75.9, 0.026)	
PSU SCIPUFF	Mass Ratio	0.203	Sources (-0.5, -0.011, 0.035)		
ENSCO 2	Mass Ratio	0.201	Puff Real (0.37, 1.3, 0)	Sources (-0.32, -0.93, 0)	
ENSCO 1	Mass Ratio	0.15	Puff Real (0.398, 14.64, 0)		
Buffalo GA	Mass Ratio	0.125	Sources (-0.365, -2.376, 0)		
Aerodyne	Mass Ratio	0.096	Puff Real (0.262, 0.852, 0.006)	Sensors (-0.212, -0.089, 0.026)	
NCAR Phase I	Mass Ratio	0			
NCAR Variation	Mass Ratio	0			
PSU Gaussian	Mass Ratio	0			
SAGE Mgt August	Mass Ratio	0			
Boise State	Mass Ratio		NO DATA		
PSU MEFA	Mass Ratio		NO DATA		
model	dependent	R2	significant factor	significant factor	significant factor
DYSTL	Mean	0.641	Puff Real (-0.807, -1.23, 0)		
NCAR Phase I	Mean	0.266	Sources (0.534, 0.09, 0.001)		
NCAR Variation	Mean	0.204	Sources (0.475, 0.09, 0.003)		
ENSCO 3	Mean	0.101	Sources (-0.35, -0.03, 0.023)		
SAGE Mgt August	Mean	0.083	Sources (0.303, 0.204, 0.002)		
ENSCO 1	Mean	0.043	Met Num (0.228, 0.009, 0.021)		
Aerodyne	Mean	0.033	Sensors (-0.206, -0.003, 0.036)		
Boise State	Mean	0			
Buffalo GA	Mean	0			
Buffalo SA	Mean	0			
ENSCO 2	Mean	0			
PSU Gaussian	Mean	0			
PSU MEFA	Mean	0			
PSU SCIPUFF	Mean	0			

Two types of regression coefficients are standardized and unstandardized. The former refers to the regression coefficients obtained after transforming all data so that the dependent variable and all the independent variables have a mean of zero and a standard deviation of 1.0. In some sense, this treats all data as being on an equal footing. The unstandardized coefficients are the result of performing regression without this transformation. For each model listed in Tables F-1 and F-2, both types of coefficients appear in parentheses after each independent variable that was selected by the regression process. The level of significance or, more technically, the "p-value" – that is the probability of the same or a more extreme outcome under the null hypothesis that this coefficient was zero – also appears in the parentheses after the coefficient. Models with gray backgrounds in Tables F-1 and F-2 are those for which there were no data or for which regression was not significant.

References

F-1. Draper, N. and H. Smith, *Applied Regression Analysis*, Wiley, 1966.

F-2. Seber, G., *Linear Regression Analysis*, Wiley, 1977.

F-3. SPSS (2006) *SPSS Base 15.0 User's Guide*, SPSS Inc., Chicago.

Appendix G
"Cross-Term" Regression Results Tables

Early in this study, we conducted analyses of variance (ANOVA) of both the mass estimation and miss distance predictions with the intent of gaining insight into which of the many factors that composed the various models had a significant effect on their outcomes. Results of the ANOVA indicated that, in certain cases, two-way interactions between factors (independent variables) were significant. With these results as motivation, we reformulated the regression equations used in the previous section to include second-order terms, such as the product of the number of sensors and the number of sources. In certain cases, this required coding categorical variables, such as diurnal conditions, as scalar quantities (e.g., assigning the value 1 to daytime and -1 to nighttime). Thus, instead of attempting to "fit" outcomes to linear functions of several variables, we attempted to model outcomes as second-order polynomials in several variables.

We then proceeded with stepwise regression, recorded the resulting adjusted R^2, and listed the significant variables and significant products in the tables below.

Table G-1. Stepwise Regression Results for "Mean Offset" Independent Variable

Model	Dependent Variable	Crossed Adjusted R2	Linear Adjusted R2	Check	Significant Factors	Significant Factors	Significant Factors
DSTL	Mean Offset	0.758	0.641	ok	Sources X Puff Real (-0.875, -0.434, 0.001)		
NCAR Phase 1	Mean Offset	0.434	0.266	ok	Sources^2 (1.18, 0.042, 0.001)	Sources X Sensors (-1.05, -0.01, 0.004)	Sensors^2 (0.603, 0.01, 0.027)
NCAR Variational	Mean Offset	0.234	0.204	ok	Sources^2 (0.504, 0.02, 0.001)		
Ensco 3	Mean Offset	0.173	0.101	ok	Sources (-1.805, -0.152, 0.014)	Sources^2 (1.486, 0.026, 0.04)	
Sage-Mgt	Mean Offset	0.085	0.083	ok	Sources^2 (0.306, 0.044, 0.002)		
Ensco 1	Mean Offset	0.08	0.043	ok	Met Num (0.230, 0.009, 0.018)	Sensors X Diurnal (-0.216, -0.001, 0.026)	
Buffalo SA	Mean Offset	0.043	0	ok	Puff Real X Met (-0.238, -0.062, 0.047)		
Aerodyne	Mean Offset	0.033	0.033	ok	Sensors (-0.206, -.003, 0.036)		
Boise State	Mean Offset	0	0	ok			
Buffalo GA	Mean Offset	0	0	ok			
Ensco 2	Mean Offset	0	0	ok			
PSU Gaussian	Mean Offset	0	0	ok			
PSU MEFA	Mean Offset	0	0	ok			
PSU SciPuff	Mean Offset	0	0	ok			

Table G-2. Stepwise Regression Results for "Mass Ratio" Independent Variable

Model	Dependent Variable	Crossed adjusted R^2	Linear adjusted R^2	Check	Significant factor 1	Significant factor 2
ENSCO 3	Mass Ratio	0.507	0.379	ok	Puff Real X MET (-1.13, -5.52, 0.001)	Sources X Puff Real (-0.646, -1.38, 0.011)
DSTL	Mass Ratio	0.475	0.254	ok	Sensors X Puff Real (-0.768, -30.2, 0.001)	Diurnal X Puff Real (-0.399, -192.3, 0.006)
Buffalo SA	Mass Ratio	0.392	0.273	ok	Sources X MET (-1.268, -1.603, 0.002)	MET Num (1.096, 2.94, 0)
ENSCO 2	Mass Ratio	0.37	0.201	ok	Puff Real (0.702, 2.43, 0.001)	Sources X Puff Real (-0.414, -0.632, 0.02)
ENSCO 1	Mass Ratio	0.307	0.15	ok	Sensors X Puff Real (0.541, 1.648, 0.001)	Sensors ^2 (0.438, 0.103, 0.001)
NCAR Phase I	Mass Ratio	0.269	0	ok	Puff Real ^2 (-0.537, -0.396, 0.001)	
PSU Gaussian	Mass Ratio	0.264	0	ok	Puff Real ^2 (0.528, 3.83, 0.001)	
PSU SCIPUFF	Mass Ratio	0.217	0.203	ok	Puff Real ^2 (-0.513, -0.030, 0.029)	
Buffalo GA	Mass Ratio	0.171	0.125	ok	Sources (-0.362,-2.35 , 0.001)	Diurnal X Puff Real (-0.232, -1.68, 0.011)
Aerodyne	Mass Ratio	0.096	0.096	ok	Puff Real (0.262, 0.85, 0.006)	Sensors^2 (-0.212, -0.004, 0.026)
NCAR Variational	Mass Ratio	0	0	ok		
SAGE-Mgt	Mass Ratio	0	0	ok		
Boise State	Mass Ratio	No data	No data	ok		
PSU MEFA	Mass Ratio	No data	No data	ok		

Model	Dependent Variable	Crossed adjusted R^2	Linear adjusted R^2	Check	Significant factor 3	Significant factor 4
ENSCO 3	Mass Ratio	0.507	0.379	ok	Sources (-0.459, -1.96, 0.001)	Puff Real ^2 (0.246, 2.56, 0.040)
DSTL	Mass Ratio	0.475	0.254	ok	Sources X Sensors (-0.305, -3.392, 0.034)	
Buffalo SA	Mass Ratio	0.392	0.273	ok	Sources (-1.090, -2.26, 0.001)	Diurnal X Puff Real (-0.256, -0.668, 0.009)
ENSCO 2	Mass Ratio	0.37	0.201	ok	Sources (-0.378, -1.096, 0.001)	Puff Real ^2 (0.365. 2.108, 0.001)
ENSCO 1	Mass Ratio	0.307	0.15	ok	Puff Real ^2 (0.299, 18.201, 0.001)	Sources X Sensors (-0.293, -0.524, 0.018)
NCAR Phase I	Mass Ratio	0.269	0	ok		
PSU Gaussian	Mass Ratio	0.264	0	ok		
PSU SCIPUFF	Mass Ratio	0.217	0.203	ok		
Buffalo GA	Mass Ratio	0.171	0.125	ok		
Aerodyne	Mass Ratio	0.096	0.096	ok		
NCAR Variational	Mass Ratio	0	0	ok		
SAGE-Mgt	Mass Ratio	0	0	ok		
Boise State	Mass Ratio	No data	No data	ok		
PSU MEFA	Mass Ratio	No data	No data	ok		

Model	Dependent Variable	Crossed adjusted R^2	Linear adjusted R^2	Check	Significant factor 5	Significant factor 6
ENSCO 3	Mass Ratio	0.507	0.379	ok		
DSTL	Mass Ratio	0.475	0.254	ok		
Buffalo SA	Mass Ratio	0.392	0.273	ok		
ENSCO 2	Mass Ratio	0.37	0.201	ok	Diurnal X Puff Real (-0.229, -0.745, 0.14)	Sensors ^2 (0.213, 0.005, 0.009)
ENSCO 1	Mass Ratio	0.307	0.15	ok		
NCAR Phase I	Mass Ratio	0.269	0	ok		
PSU Gaussian	Mass Ratio	0.264	0	ok		
PSU SCIPUFF	Mass Ratio	0.217	0.203	ok		
Buffalo GA	Mass Ratio	0.171	0.125	ok		
Aerodyne	Mass Ratio	0.096	0.096	ok		
NCAR Variational	Mass Ratio	0	0	ok		
SAGE-Mgt	Mass Ratio	0	0	ok		
Boise State	Mass Ratio	No data	No data	ok		
PSU MEFA	Mass Ratio	No data	No data	ok		

For the majority of cases, the cross-term regression results are completely consistent with the linear regression results presented in the main body of the report – when cross-term factor is determined to be significant by the regression, then either (or both) of the two factors is/are determined to be significant by no cross-term regression. The main exception for this is PSU SCIPUFF for "Mass Ratio" dependent variable. Further examination of the PSU SCIPUFF predictions reveals that the algorithms performed rather poorly in terms of predicting the mass of the release. This is especially true for releases when a high amount of material was released (e.g., continuous releases or multiple realizations of instantaneous releases). Both the "Puff Real^2" and the "Sources" independent variables have a strong correlation with the total amount of material released.

Appendix H
Task Order Extract

(with most pertinent section in red font)

DC-1-2607

TITLE: Support for DTRA in the Validation Analysis of Hazardous Material Assessment Models

This task order is for work being performed by the Institute for Defense Analyses (IDA) under Contract Number W91WAW-09-C-0003 (see paragraph 9e) for the Defense Threat Reduction Agency (DTRA).

1. BACKGROUND:

The DTRA/Joint Science and Technology Office (JSTO) Verification and Validation (V&V) Program represents ongoing activities performed in parallel with development of all predictive codes in support of hazardous material transport and dispersion prediction. One element of V&V is to perform code-on-code comparisons. In this strategy, each code receives the same input. In this manner, differences in the output predictions can lead to the identification of software bugs, or help to assess technical strengths and weaknesses of component algorithms within each code. In addition, a certain amount of credibility for both models is achieved when their predictions agree. When the inputs are simple, such as for fixed winds and simple terrain, the predictions tend to be dominated by the dispersion algorithms. Comparisons at this level of complexity are important to establish fundamental dispersion algorithm veracity, and to help discover software bugs. As more complex terrain, urban landscapes, and weather are included as inputs, the number of physical processes responsible for transport and dispersion increases and the predictions become the result of many interdependent algorithm calculations.

It is very difficult to separate meteorological uncertainty from the transport and dispersion model accuracy when comparing predictions to field-trial validation quality or real-world data. The validation challenge is to assess whether a model performs well over different field trials, and ultimately reflects real-world phenomena. Some codes perform better under certain conditions and specific scenarios. Hazard prediction models are generally developed for a range of user communities and applications. Each user community has a different set of requirements. Thus, the corresponding hazard models tend to be optimized for specific applications. The process of validating a model should be couched in terms of end-user requirements, where feasible.

Several aspects of hazard prediction modeling are the subject of current improvement programs:

1. Algorithms to estimate source term parameters (e.g., location, time, and amount) from sparse observations are also being developed. Such "sensor data fusion" tools are expected to improve hazard predictions in scenarios where the release is covert or accidental. Field experiments have been conducted, and are being designed, to aid in the evaluation of urban (including within a building) and sensor data fusion models. These evaluations are crucial to the overall management of these programs.

2. Because of prohibitive cost of field trials, there is a program to develop realistic synthetic environments that would allow virtual testing and validation of CBRN sensors and models. These virtual environments could also be used for CONOPS development. Different sub-modules of these simulators should account for all potential environmental aspects that are needed for satisfactory validation of sensors and models including meteorology, atmospheric backgrounds, and simulated threat. Since these complex systems purport to simulate "reality", a rigorous validation of subcomponents is needed.

3. Complexities associated with the urban environment are being addressed via an urban transport and dispersion program. Codes varying from empirical (wind tunnel-based) to computational fluid dynamics-based are being considered to address the complex flows associated with an urban environment. As they become mature (and validated), tools to address the infiltration, exfiltration, and flow within buildings and other complex structures are also being considered for inclusion within hazard prediction models.

2. OBJECTIVE:

IDA will conduct independent analyses and special studies associated with verification, validation, and evaluation of the suite of models associated with the Hazard Assessment. IDA will support development of user-oriented performance MOEs using field trial data sets and will coordinate scenario definition and arbitration for code-on-code V&V activities.

The objectives of these analysis and coordination are (1) to ensure that a consistent analysis approach is used when comparing model predictions, and to assist DTRA in the implementation of code-on-code analysis, comparisons, and interpretation; and (2) to define measures of effectiveness in terms of user-specific objectives and applications.

The scope of this effort may be expanded to other programs as directed by DTRA.

3. STATEMENT OF WORK:

As required by DTRA technical representatives, IDA will perform the following tasks:

Sensor Data Fusion (SDF) Related Studies. IDA will provide technical and analytical support associated with the initial incorporation of SDF algorithms into hazard prediction tools and products. In order to support credible quantitative assessments of this emerging technology area, new analytical techniques and procedures/protocols are required. IDA will conduct independent comparative studies of different SDF algorithms using the data collected during Fusion Field Trial 2007 (FFT 07). Phase I of this investigation included 104 cases created using FFT 07 field trial data were distributed to eight organizations in September, 2008. Last set of predictions were received in August, 2009 when Phase I of the exercise was "officially" closed. FY10 work will include analysis and inter-comparison of the fourteen sets of predictions that were provided by different STE algorithm developers with results summarized in IDA document expected in spring FY10. Additionally, a Phase II of the exercises is planned to commence in FY10. IDA will be responsible for preparation of test cases for which predictions will be sought, overall coordination among exercise participants and final analysis and adjudication of the results.

VTHREAT Validation Analyses. As directed by the sponsor, IDA will assist with the validation of the VTHREAT synthetic environment being developed by the National Center for Atmospheric Research (NCAR). This work will be performed in close coordination and collaboration with the developer of VTHREAT. In FY09, a preliminary analysis using limited data supplied to IDA by NCAR was performed to test methodology and initial results were briefed to NCAR and sponsor. We're planning to expand this effort in FY10 to include: a) additional data supplied by NCAR and b) timely feedback provided back to NCAR so that additional improvements could be implemented in VTHREAT.

Building Interior T&D Model Validation. JEM Increment 3 includes a requirement to include building interior T&D. IDA in coordination with NSWCDD will provide support to validation of building interior T&D modeling to be included in JEM. This work could involve either comparison of T&D models against available field trial data or code-on-code comparisons.

V&V of Urban Dispersion Modeling. Complex Urban dispersion modeling is an active area where T&D modeling improvements are sought. To that effect, IDA will continue V&V studies involving comparisons of urban T&D with field trials. IDA is exploring possibility of using Urban Dispersion Program (UDP) field trials that included two sets of tracer releases in NYC for validation of UDM and Micro-SWIFT/Micro-SPRAY urban dispersion codes. Additionally, IDA will continue efforts supporting validation of the latest version of Micro-SWIFT/Micro-SPRAY with Urban 2000 and Joint Urban 2003 field trials data.

Meteorological Studies Associated with FFT 07 Data. As directed by sponsor, IDA will conduct studies and analyses of a vast meteorological dataset collected by a dense grid of PWIDS during FFT 07 Field Trials. This task will greatly benefit from an expected close collaboration and coordination with Meteorology Division of Dugway Proving Ground and DTRA meteorologists.

a. As a part of the all of the above subtasks, IDA will communicate, via conference papers and/or posters, working group discussions, and IDA papers, the more important applications of the MOE and any progress toward the creation of "demonstration" validations. In addition, IDA should create descriptions of its efforts, where appropriate (and approved by DTRA), that are suitable for publication in peer-reviewed journals. IDA will actively participate in working groups (e.g., Sensor Data Fusion), Science Teams for potential upcoming experiments and Technical Panel 9 as directed by DTRA. As required, IDA will provide independent reviews (e.g., of proposals or of JSTO-funded programs) and may assist DTRA with international collaborative comparative efforts (e.g., with Israel or UK).

4. CORE STATEMENT:

This research is consistent with IDA's mission in that it will support specific analytical requirements of the sponsor and will assist the sponsor with planning efforts. Accomplishment of this task order requires an organization with experience in operationally oriented issues from a joint and combined perspective, which IDA, a Federally Funded Research and Development Center, is able to provide. It draws upon IDA's core competencies in Systems Evaluations and Operational Test and Evaluation. Performance of this task order will benefit from and contribute to the long-term continuity of IDA's research program.

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**		
1. REPORT DATE (DD-MM-YY) July 2012	2. REPORT TYPE Final	3. DATES COVERED (FROM – TO) September 2008 – September 2010
4. TITLE AND SUBTITLE Comparative Investigation of Source Term Estimation Algorithms for Transport and Dispersion Modeling Based on Limited Number of Sensor Readings		5A. CONTRACT NO. DASW01-04-C-0003
		5B. GRANT NO.
		5C. PROGRAM ELEMENT NO(S).
6. AUTHOR(S) Platt, N; DeRiggi, D.F.		5D. PROJECT NO.
		5E. TASK NO. DC-1-2607 AND DC-1-2615
		5F. WORK UNIT NO.
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882		8. PERFORMING ORGANIZATION REPORT NO. IDA Document D-4048
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Threat Reduction Agency Joint Science & Technology Office for Chemical & Biological Defense 8725 John J. Kingman Road Fort Belvoir, VA 22060-6201		10. SPONSOR'S / MONITOR'S ACRONYM(S) DTRA-JSTO
		11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

Dr. Nathan Platt, Project Leader

14. ABSTRACT

The release of hazardous materials into the atmosphere on the battlefield or in populated areas must be considered for future scenarios. Given a warning based on detections at a few sensors, it should be useful to rapidly provide an estimate of the location, time of release, and amount of material released. Such information could lead to refined predictions of the hazardous area and support follow-on actions to investigate the cause and nature of the hazardous release. In September 2007, a short-range test – Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07) – designed to collect data to support development of prototype source term estimation (STE) algorithms was conducted. A comparative investigation of STE algorithms began in 2008. First, a subset of sensor data from selected FFT 07 trials was provided to participating algorithm developers. Next, developers provided "blind" STE predictions that were then independently compared to parameters of the actual release. Eight STE algorithm developers participated in this exercise. Fourteen full and partial sets of predictions were received with some exercise participants providing multiple sets of predictions based on different algorithms they have been developing. This evaluation considered several variables that might influence results including the number of sensors (4 vs. 16), the release type (instantaneous vs. continuous), the time of the release (day vs. night), meteorological inputs ("research-grade" inputs vs. "simulated" operational inputs), and the number of sources (single vs. double vs. triple vs. quad releases). The results of these analyses are used to ascertain trends among different sets of STE predictions and are presented in this paper.

15. SUBJECT TERMS

sensor data fusion; source term estimation; FUSION field trial 2007; FFT 07;

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19A. NAME OF RESPONSIBLE PERSON Dr. John Hannan
A. REPORT Unclassified	B. ABSTRACT Unclassified	C. THIS PAGE Unclassified	Unlimited	80	19B. TELEPHONE NUMBER (INCLUDE AREA CODE) 703-767-3286